EVOLUTION OF THE UNNECESSARY: INVESTIGATING HOW FMET BECAME CENTRAL IN BACTERIAL TRANSLATION INITIATION

A thesis submitted in partial fulfilment of the requirements for the degree of

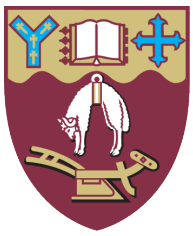
Doctor of Philosophy
in Cellular and Molecular Biology

at the

University of Canterbury

by

Ryan J. Catchpole



2014

Table of contents

Acknowledgements

I wish to thank everyone who has made this thesis possible.

In particular, thank you to Jack and Ant for your support and guidance, and for allowing me to work on this crazy idea.

To all of the lab members and staff who helped me, listened to me and entertained me over the past five years, thank you – you are too numerous to list, but your support has meant so much in these [literally] shaky times.

To Mum and Dad, thank you for sticking by me through this long process. Your reassurance and encouragement makes every hill surmountable.

And Tom. Your love and support means more than you can imagine.

Abstract

All bacteria initiate translation using formylated methionine, yet directly after translation, the formyl-group is removed. This sequence of addition and removal appears futile, yet every sequenced bacterial genome encodes the enzymes for formylation and deformylation, suggesting this process is essential. Puzzlingly, the process is absent from both Archaea and Eukaryotes, and moreover, bacterial mutants lacking both the formylase and deformylase activities are viable, albeit with a diminished growth rate.

We created an *Escherichia coli* strain devoid of formylase and deformylase activity. This strain was then allowed to evolve over 1500 generations whereupon it reached wild-type growth rate, demonstrating that formylation can be completely dispensed with. This raises an additional question: if the formylation cycle is unnecessary, how did it emerge and why has it persisted?

Our results show that the formylation-deformylation cycle could have evolved as a toxin-antitoxin pair (TA) with post-segregational killing (PSK) activity. TAs 'addict' cells to the plasmids that carry them by inducing PSK. We measured the stability of formylase-deformylase encoding plasmids and their ability to elicit PSK in our evolved *E. coli* strain. We report several lines of evidence consistent with the formylation-cycle having evolved from a plasmid-borne PSK element: 1) in the absence of deformylation, formyl-methionine on proteins is cytotoxic in bacteria 2) deformylation relieves the cytotoxicity of formyl-methionine, 3) the loss of a plasmid containing formylase and deformylase genes from evolved cells results in cessation of growth – a standard PSK phenotype.

In addition, we introduced the *E. coli* formylase and deformylase genes into yeast and demonstrate that Met-tRNA formylation is not lethal, even in the absence of deformylation. This suggests PSK would be ineffectual in yeast, accounting for the absence of formylation from eukaryotic cytoplasmic translation.

We also report the presence of formylase and deformylase genes in the two representative members of the archaeal *Methanocopusculum* genus. Moreover, we demonstrate that these genes have been acquired by a recent horizontal gene transfer from bacteria.

Our results indicate that formylmethionine use in bacteria evolved, not through a direct functional benefit to cells, but through competition between infectious genetic elements.

Chapter 1
Introduction

Translation is the process by which all cells synthesise polypeptides and proteins. The conservation of ribosome-mediated translation implies that it was present in the last universal common ancestor of life (Hoeppner et al., 2012; Hsiao et al., 2009; Petrov et al., 2014). As a result, translation proceeds in a similar way in eukaryotes, archaea and bacteria, able to be separated in to three major phases: translation initiation, elongation and termination (Ramakrishnan, 2002). The initiation phase involves assembly of initiation factors and ribosomal subunits on a messenger RNA molecule, and detection of the start codon of the transcribed gene by the initiator tRNA (Kozak, 1999). The elongation phase involves the movement of the ribosome along the mRNA strand, and the use of elongator tRNAs and aminoacyl-tRNA synthetases to ensure fidelity between the specified codons and amino acids tethered through the peptidyl-transfer reaction, producing a nascent polypeptide chain (Schmeing and Ramakrishnan, 2009). The termination stage involves recognition of the stop codon on the mRNA by release factors, followed by release of the polypeptide chain and disassembly of the ribosome (Nakamura et al., 1996).

Although the broad features of translation are shared among eukaryotes, archaea and bacteria, there are some important differences in the way each stage is carried out across the three domains. Some oddities in translation are highly conserved within one or two domains, but not universally across the tree of life. This raises the question: why are certain features of translation seemingly important in some organisms, but unnecessary in others? One such feature is the use of formylmethionine in bacterial translation initiation. Although highly conserved in

bacteria, formylmethionine is not used in archaeal or eukaryotic cytoplasmic translation (Yuan 2001). The initiation of translation in each domain, including the role of formylmethionine, is described below in more detail.

Translation initiation in Bacteria

Translation in bacteria utilizes three initiation factors: IF1, IF2 and IF3. Each of these initiation factors is able to bind the 30S ribosomal subunit independently, with no requirement for ordered binding (Milón et al., 2012). Despite this, it appears there is a kinetically favoured binding order with the binding of each initiation factor increasing the affinity of the pre-initiation complex for subsequent initiation factors. The binding of IF3 to the 30s ribosomal subunit is most kinetically favourable, and thus, likely to be the first binding event (Milón et al., 2012). The C-terminal domain of IF3 associates with the subunit interface of the 30S ribosomal subunit and prevents aberrant assembly of 70S ribosomes by obstructing binding of the 50S ribosomal subunit (Dallas and Noller, 2001). IF2 is a GTPase that binds to the 30S-IF3 pre-initiation complex forming IF2·GTP soon after the binding of IF3. This association stabilises the interaction between IF3 and the 30S subunit and increases the affinity of the pre-initiation complex for IF1. Additionally, the binding of IF1 strengthens the association of both IF2·GTP and IF3 with the 30S subunit, further stabilising the pre-initiation complex (Figure 1a) (Milón and Rodnina, 2012).

The binding of mRNA to the 30S ribosomal subunit is determined primarily by the strength of the interaction between the ribosome-binding site (Shine-Dalgarno sequence) on the mRNA and the 3' sequence of the 16S rRNA (anti-Shine-Dalgarno sequence) in the 30S subunit (Studer and Joseph, 2006). As this binding is independent of initiation factors, it can occur at any time during the assembly of the 30S-IF1-IF2·GTP-IF3 pre-initiation complex (Figure 1b) (Milón et al., 2012).

The final component required for formation of the complete 30S initiation complex is the formylmethionyl-initiator tRNA (fMet-$tRNA_i^{fMet}$). fMet-$tRNA_i^{fMet}$ is recruited to the initiation complex by interaction of the formylmethionine amino acid with the C-terminal domain of IF2 (Milón et al., 2010; Spurio et al., 2000). This interaction positions the anticodon of fMet-$tRNA_i^{fMet}$ in the ribosomal P-site (Gabashvili et al., 2000). The start site of mRNA is also positioned close to the ribosomal P-site by the Shine-Dalgarno:anti-Shine-Dalgarno interactions of the 16S rRNA. Recognition of the start-codon requires interaction between the AUG codon of the mRNA and the CAU anti-codon of the fMet-$tRNA_i^{fMet}$. This interaction causes conformational changes in the initiation complex that further stabilise the binding of IF1, IF2·GTP, mRNA and fMet-$tRNA_i^{fMet}$ to the 30S subunit, while destabilising the binding of IF3 (Figure 1c) (Milón et al., 2012). Consequently, IF3 is readily displaced from the pre-initiation complex by the 50S ribosomal subunit (Fabbretti et al., 2007; Pon and Gualerzi, 1986). The GTPase-associated center of the 50S ribosomal subunit interacts with IF2·GTP to promote GTP hydrolysis (Qin et al., 2009) (Figure 1d). The resulting IF2·GDP+P_i has a lower affinity for the initiation complex than IF2·GTP, and is released along with IF1 (Figure 1e) (Antoun et al., 2003). The mature 70S ribosome is now able to begin peptide synthesis.

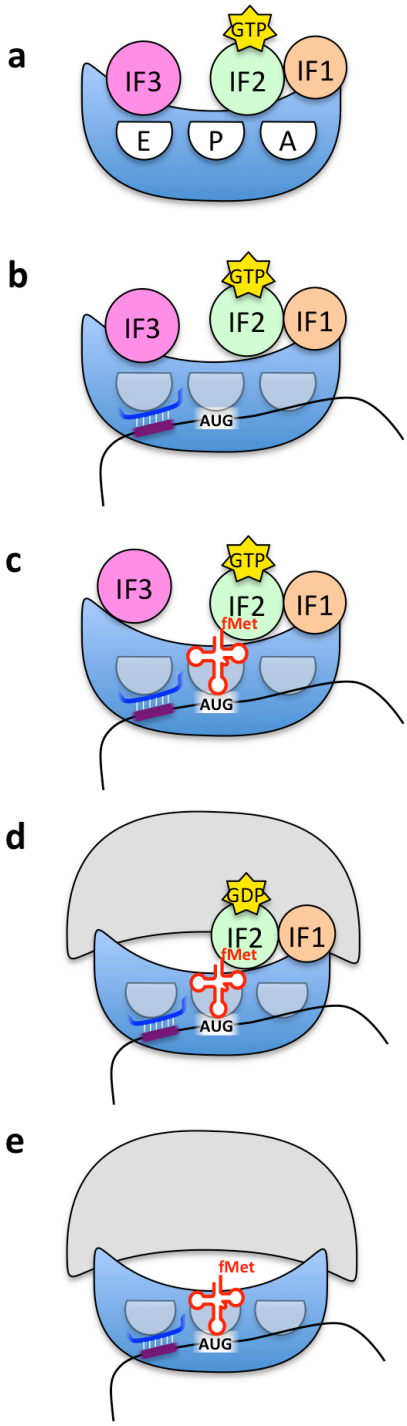


Figure 1 | Bacterial translation initiation. a. IF1, IF2·GTP and IF3 bind to the 30S subunit (blue). Ribosomal E-, P- and A- sites are labelled. **b.** mRNA binds to the 30S subunit via Shine-Dalgarno (purple) and anti-Shine-Dalgarno (cobalt) interactions. This positions AUG in ribosomal P-site. **c.** fMet-tRNA$_i^{fMet}$ (red) binds to IF2 in the P-site, interacting with the AUG start codon. IF3 is destabilised. **d.** 50S subunit (grey) binds to 30S subunit. IF3 dissociates. GAC promotes GTP hydrolysis on IF2. **e.** IF1 and IF2·GDP dissociate from ribosome. Translation elongation can begin.

Translation initiation in Eukaryotes

In contrast to bacteria, translation initiation in eukaryotes proceeds with a non-formylated methionyl-$tRNA_i$. However, eukaryotic translation has its own idiosyncrasies, involves a greater number of initiation factors, and a more complex network of interactions than in bacteria. The assembly of two protein:RNA complexes is required before translation can begin: The activated mRNA complex (containing the mRNA to be translated) and the pre-initiation complex (containing the 40S ribosomal subunit).

The activated mRNA complex requires a mature mRNA with two important post-transcriptional modifications: a 7-methylguanosine nucleotide at the 5' position (m^7G cap) and a polyadenylated 3′ end (poly-A tail). These modifications are subject to binding by translation initiation factors. Eukaryotic translation initiation factor 4E (eIF4E) binds to the m^7G cap and recruits the scaffold protein eIF4G. Both eIF4E and eIF4G bind the ATP-dependent RNA helicase eIF4A. The helicase activity of eIF4A is stimulated by eIF4G (Andreou and Klostermeier, 2014; Oberer et al., 2005) (as well as other eIF4 initiation factors (Rogers et al., 2001)) and removes secondary structure from the 5′-untranslated region (5′-UTR) of the mRNA. The m^7G bound eIF4G-eIF4E-eIF4A complex is named eIF4F. Poly-A binding proteins (PABP) associate with the poly-A tail of the mRNA, and also with eIF4F, bringing together the 3′ and 5′ ends of the mRNA molecule. (Figure 2a)

The pre-initiation complex (PIC) requires assembly of initiation factors to the 40S ribosomal subunit, as well as formation of the Met-$tRNA_i^{Met}$:eIF2·GTP moiety known as the ternary complex (Figure 2a). The initiation factors eIF5 and eIF3 bind at the P- and E-sites, respectively. The binding of eIF1 to the P-site of the 40S ribosomal subunit and eIF1A to the A-site promote binding of the ternary complex to the P-site, and open the mRNA binding channel in

preparation for binding of the activated mRNA complex (Passmore et al., 2007; Weisser et al., 2013). The presence of eIF1 prevents Met-tRNA$_i^{Met}$ from fully occupying the P-site. This complete PIC is then able to bind the activated mRNA complex (Figure 2b).

Interactions between eIF3 of the PIC and eIF4F of the mRNA complex bring these two large protein:RNA complexes together to form a scanning complex (Hinnebusch, 2006). An interaction between eIF5 and eIF2·GTP promotes hydrolysis of GTP to GDP+P$_i$, however release of P$_i$ is prevented by eIF1 (Figure 2c). The PIC scans the mRNA until an AUG start codon is identified with a suitable surrounding sequence context (Kozak sequence (Kozak, 1987)). During scanning, eIF4A disrupts secondary structure in the mRNA and eIF1A ensures a stable association of the mRNA with the PIC. The start codon is recognised by the interaction between the CAU anticodon of Met-tRNA$_i^{Met}$, and the AUG codon of the mRNA. This interaction displaces eIF1 from the P-site, allowing accommodation of Met-tRNA$_i^{Met}$ into the P-site, and allowing P$_i$ release from eIF2 (Figure 2d). eIF5 and eIF2·GDP are then released from the initiation complex allowing eIF5B·GTP to bind to Met-tRNA$_i^{Met}$ (Figure 2e). eIF5B promotes association of the 60S ribosomal subunit with the 40S subunit to form an 80S pre-initiation complex. The association of the 60S subunit activates GTP hydrolysis by eIF5B and the release of eIF5B·GDP, eIF1A and eIF3 (Figure 2f). The 80S ribosome is now able to carry out translation.

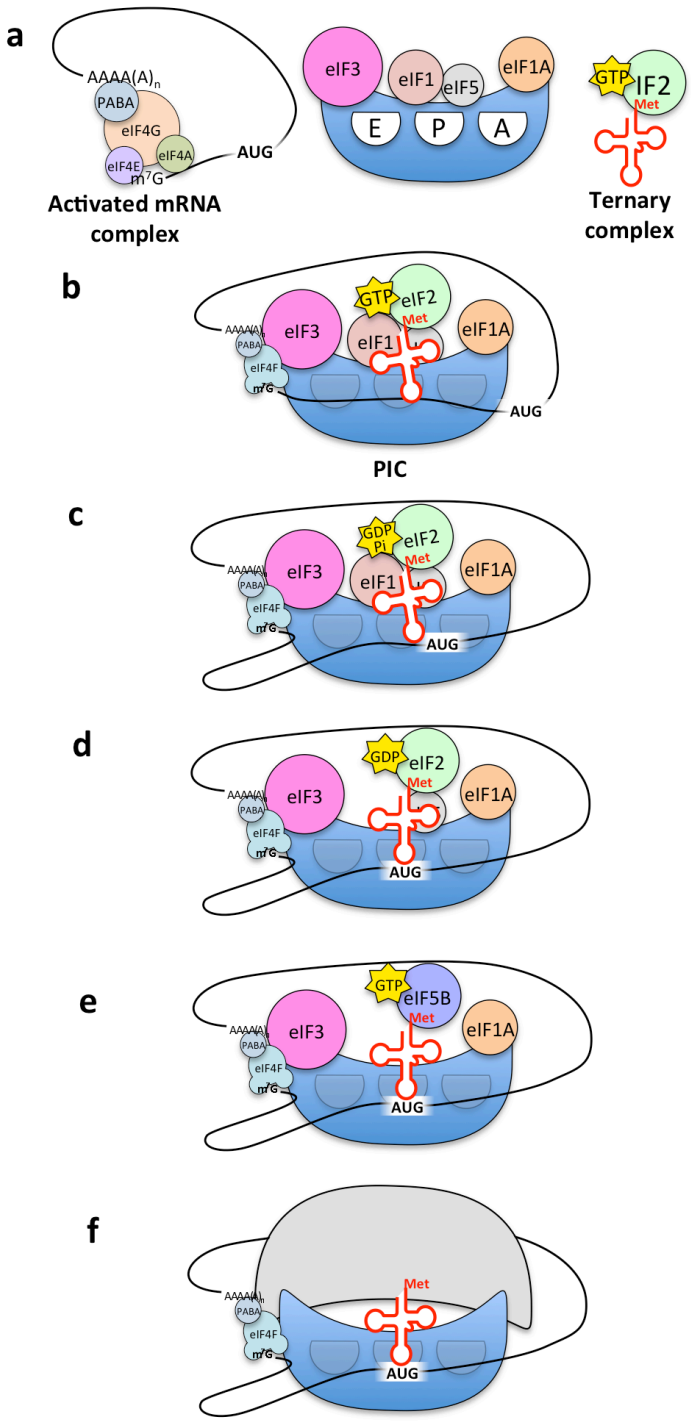


Figure 2 | Eukaryotic translation initiation. a. eIF4E, eIF4A and eIF4G bind to 5′ of mRNA, PABA binds to 3′. eIF1, eIF1A, eIF5 and eIF3 bind to 40S ribosomal subunit (blue). IF2·GTP binds to non-formylated Met-tRNA$_i^{Met}$. **b.** eIF3:eIF4F interactions promote binding of mRNA to ribosome. Ternary complex binds at P-site, partially obstructed by eIF1. **c.** GTP hydrolysis in eIF2 initiates scanning of mRNA. **d.** Start codon identification displaces eIF1, allowing P_i release from eIF2. eIF2·GDP and eIF5 are released. **e.** eIF5B·GTP binds to Met-tRNA$_i^{Met}$. **f.** GTP hydrolysis by eIF5B promotes 60S subunit (grey) joining and release of remaining initiation factors eIF5B·GDP, eIF1A and eIF3. Ribosome can begin translation.

Translation initiation in Archaea

How the translation reactions proceed in archaea are not as thoroughly characterised as they are for bacteria or eukaryotes, however translation initiation is known to proceed in a manner similar to bacteria, using initiation factors similar to eukaryotes. Similar to that observed in eukaryotes, archaeal initiation factors aIF1 and aIF1A bind to the 30S ribosomal subunit, promoting the binding of aIF2·GTP to the ribosomal P-site (Hasenöhrl et al., 2006) (Figure 3a). Archaeal initiation factor 2 (aIF2) is homologous to eIF2, showing little similarity to bacterial IF2 (Kyrpides 1998). Despite this, the aIF2·GTP:Met-$tRNA_i^{Met}$ ternary complex of archaea forms by binding of Met-$tRNA_i^{Met}$ to the ribosome bound aIF2·GTP (Figure 3c) as in bacteria, rather than prior to ribosome binding as in eukaryotes (Hasenöhrl et al., 2009). The presence of both aIF1 and aIF2·GTP:Met-$tRNA_i^{Met}$ on the 30S subunit promotes binding of the mRNA. As in bacteria, this binding is often directed by the interaction of an mRNA-based Shine-Dalgarno sequence with the 16S rRNA based anti-Shine-Dalgarno sequence (Dennis, 1997; Sartorius Neef and Pfeifer, 2004) (Figure 3b). The start codon is recognised by interaction between the AUG codon of the mRNA and the CAU anticodon of Met-$tRNA_i^{Met}$. This interaction promotes hydrolysis of GTP in aIF2 (Figure 3c), followed by release of aIF1A and aIF2·GDP. Archaea have a homologue of eIF5B, however it is currently unclear whether this protein is responsible for GTP-dependent ribosome assembly (Figure 3d), as in eukaryotes, or functions as an alternative aIF2, facilitating Met-$tRNA_i^{Met}$ binding to the 30S subunit.

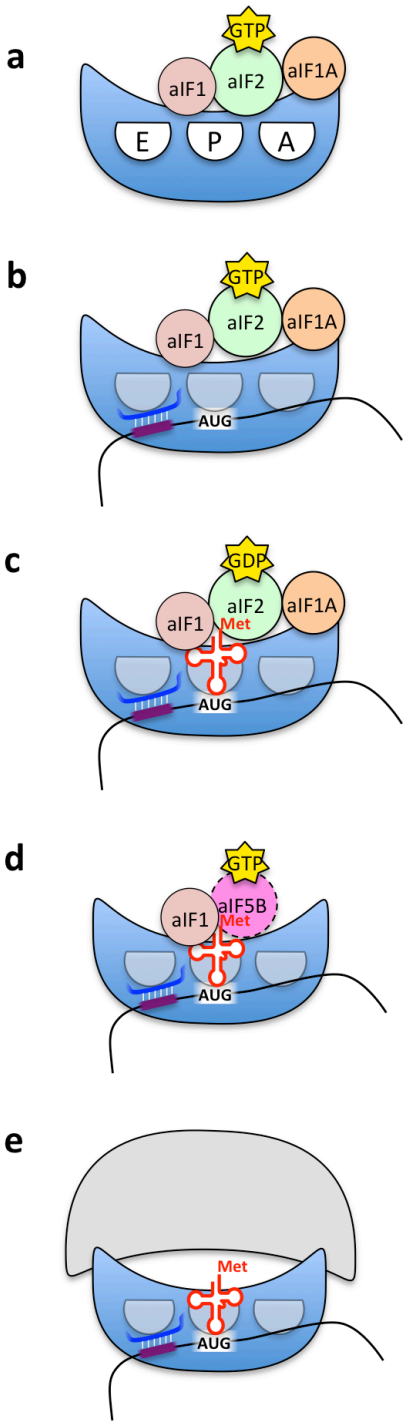


Figure 3 | Archaeal translation initiation. a. aIF1A, aIF2·GTP and aIF1 bind to the 30S subunit (blue). Ribosomal E-, P-, and A-sites are labelled. **b.** mRNA binds to the 30S subunit via Shine-Dalgarno (purple) and anti-Shine-Dalgarno (cobalt) interactions. This positions AUG in ribosomal P-site. **c.** Met-tRNA$_i^{Met}$ (red) binds to aIF2 in the P-site, interacting with the AUG start codon. GTP is hydrolysed in aIF2 promoting dissociation of aIF1A and aIF2·GDP. **d.** It is unknown whether aIF5B·GTP (pink) facilitates GTPase-dependent subunit joining. **e.** 50S subunit (grey) binds to 30S subunit. Remaining initiation factors dissociate. Ribosome can begin translation elongation.

The mechanisms presented above represent a single translation process in each domain. Subtle differences in translation initiation facilitate expression from leaderless mRNAs in all domains (Benelli et al., 2003; Li and Wang, 2004; Moll et al., 2004), internal ribosome entry sites in eukaryotes and viruses (Filbin and Kieft, 2009), and polycistronic mRNAs in bacteria (Karamyshev et al., 2004) and archaea (Kramer et al., 2014) (and rarely in eukaryotes (Gould et al., 2014)).

It is clear that many features of translation initiation are conserved across all three domains of life. Many features such as translation initiation factors 1 (bacterial IF1 is functionally homologous to a/eIF1A (Battiste et al., 2000; Kyrpides and Woese, 1998)) and 2 (bacterial IF2 is homologous to a/eIF5B and functionally analogous to a/eIF2 (Lee et al., 2002)), the use of a specialised initiator tRNA and the ribosome itself are shared between bacteria, eukaryotes and archaea likely owing to their presence in the last universal common ancestor (Fox and Naik, 2004). Despite this conservation, some features of translation initiation have clearly diverged during evolution of the three domains. This is most obvious in eukaryotes where the 5' m7G cap (as well as the associated eIF4 initiation factors) and mRNA scanning are features completely absent from both bacteria and archaea (Shuman, 2002). Additionally, the use of formylmethionyl-$tRNA_i$ is unique to bacterial translation, and to the bacterial-endosymbiont-derived organelles of eukaryotes, the mitochondria and plastids (Giglione and Meinnel, 2001).

Bacterial Met-tRNA$_i$ is formylated

All bacteria initiate translation using formylated methionine (Marcker and Sanger, 1964). In contrast, both archaeal and eukaryotic cytoplasmic translation proceeds with non-formylated methionine (Smith and Marcker, 1970; White and Bayley, 1972). Formylmethionyl-tRNA$_i$ is formed by formylation of a methionylated-initiator tRNA (Met-tRNA$_i^{fMet}$) – a reaction catalysed by the enzyme methionyl-tRNA formyltransferase (EC 2.1.2.9) using the formyl-donor 10-formyltetrahydrofolate as a cofactor (Schomburg and Stephan, 1996) (Figure 4).

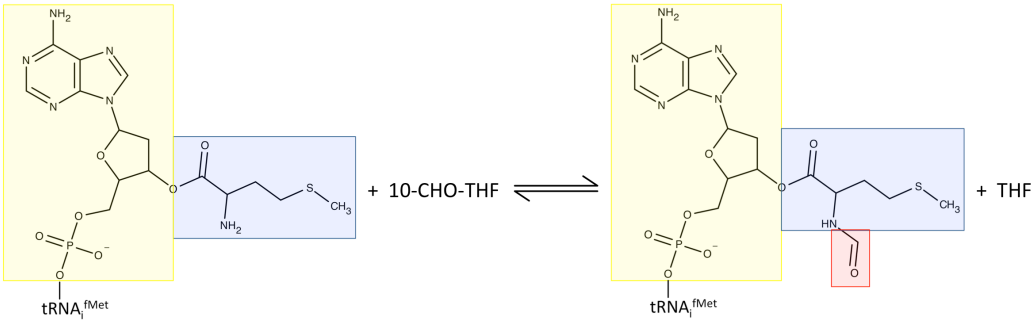


Figure 4 | Formylaton of Met-tRNA$_i^{fMet}$ by methionyl-tRNA formyltransferase. Met-tRNA$_i^{fMet}$ is formylated by transfer of the formyl-group from 10-formyltetrahydrofolate (10-CHO-THF) to the methionine residue of Met-tRNA$_i^{fMet}$. The reaction produces fMet-tRNA$_i^{fMet}$ and tetrahydrofolate (THF). The 3' nucleotide of the initiator tRNA (adenine) is highlighted in yellow and the methionine moiety highlighted in blue. The resulting formyl-group is highlighted in red.

It is important that the formylation reaction is specific to Met-$tRNA_i^{fMet}$ - formylation of the elongator tRNA, Met-$tRNA_e^{Met}$, would block the amino group and obstruct peptide bond formation. Consequently, Met-$tRNA_e^{Met}$ would be unable to participate in translation initiation, halting protein synthesis. For the process in *E. coli*, this specificity is achieved by a combination of features on both the methionyl-tRNA formyltransferase (formylase) enzyme and the Met-tRNA molecules. The active site of the formylase enzyme is located in a deep crevice in the protein structure. This crevice is longer than the single stranded acceptor stem of both Met-$tRNA_i^{fMet}$ and Met-$tRNA_e^{Met}$. As a result, binding of the methionyl moiety of Met-tRNA to the active site requires bending of the tRNA structure and disruption of the first base pair of the tRNA acceptor stem (Schmitt et al., 1998). In both *Pseudomonas aeruginosa* and *E. coli* $tRNA_i^{fMet}$, the first base pair is C_1:A_{72} (Figure 5a) – a non-Watson-Crick base pair which is easily disrupted by the formylase enzyme to facilitate entry of the methionylated acceptor stem to the active site (Newton et al., 1999c). In contrast, *E. coli* $tRNA_e^{Met}$ has a G_1:C_{72} base pair in the first position, preventing access of the methionyl moiety to the active site of formylase and decreasing formylation activity at least 495-fold (Lee et al., 1991). Additionally, the U_4:A_{69} pairing of $tRNA_e^{Met}$ is weakly inhibitory to formylation (Lee et al., 1991).

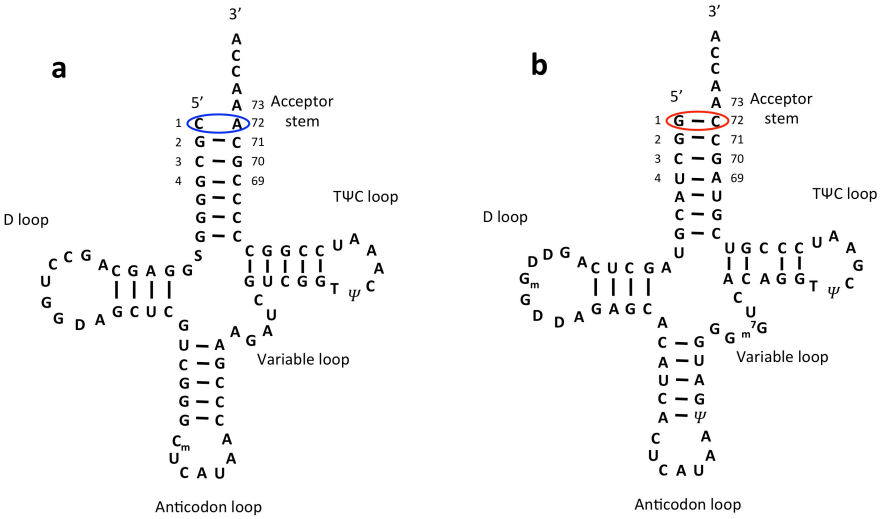


Figure 5 | *E. coli* $tRNA_i^{fMet}$ and $tRNA_e^{Met}$. a) initiator $tRNA_i^{fMet}$ from *E. coli* with the C_1:A_{72} non-Watson-Crick base pair important for Met-tRNA formylation circled in blue. Base pairs G_2:C_{71}, C_3:G_{70} and G_4:C_{69} also promote formylation activity. b) elongator $tRNA_e^{Met}$ from *E. coli* with the G_1:C_{72} base pair strongly inhibitory to formylation circled in red.

The formylmethionine moiety of fMet-$tRNA_i^{fMet}$ is a major point of recognition by IF2 during translation initiation. The *C*-terminal domain of bacterial IF2 (IF2-C2) contains a cleft which binds the formylmethionine and 3' nucleotide of the $tRNA_i^{fMet}$ acceptor stem (Guenneugues et al., 2000; Meunier et al., 2000; Szkaradkiewicz et al., 2000). This IF2-C2 cleft is able to distinguish between formylated and non-formylated Met-$tRNA_i^{fMet}$, displaying a high affinity for fMet-$tRNA_i^{fMet}$ and a very low affinity for Met-$tRNA_i^{fMet}$. In *E. coli* and *Bacillus stearothermophilus,* this differential binding can be attributed to the positive charge of the $tRNA_i^{fMet}$ binding cleft of IF2-C2 which repels the free amino group of Met-$tRNA_i^{fMet}$ through unfavourable electrostatic interactions (Steiner-Mosonyi et al., 2004). In contrast, the *N*-blocked fMet-$tRNA_i^{fMet}$ has a lower positive charge density and is able to interact with the IF2-C2 domain, promoting binding of fMet-$tRNA_i^{fMet}$ to the ribosome. Interestingly, IF2 from *P. aeruginosa* and from the mitochondria of *Saccharomyces cerevisiae* show a neutral charge in the fMet-$tRNA_i^{fMet}$ binding cleft (Steiner-Mosonyi et al., 2004). Consistent with this, both proteins have near equal affinity for both fMet-$tRNA_i^{fMet}$ and Met-$tRNA_i^{fMet}$ (Garofalo et al., 2005; Newton et al., 1999a).

Bacterial peptides are deformylated

The use of fMet-$tRNA_i^{fMet}$ as the initiating entity in translation results in the production of peptides with a formylmethionine residue at the *N*-terminus, however, only a handful of *E. coli* proteins retain this formylmethione residue (Coleman, 1992; Guan et al., 2011; Milligan and Koshland, 1990). The formyl-group is hydrolysed by the zinc-metallopeptidase enzyme, peptide deformylase (deformylase) (Adams, 1968; Meinnel and Blanquet, 1995). Deformylase activity requires a minimum of one amino acid following the formylmethionine residue, thus activity is specific towards polypeptides rather than fMet-$tRNA_i^{fMet}$ (Meinnel and Blanquet, 1995). The deformylase enzyme binds at the ribosomal

exit site and removes the formyl-moeity from the *N*-terminus of the nascent polypeptide as it emerges from the ribosome (Bingel-Erlenmeyer et al., 2008). As ribosomes are approximately 30 times more abundant than deformylase enzymes, comprehensive deformylation requires the enzyme to quickly scan ribosomes (Sandikci et al., 2013).

In addition to deformylation, the methionine residue is removed in over 80% of *E. coli* proteins (Matheson and Yaguchi, 1975). This methionine hydrolysis is essential to the function of many proteins, especially those where the *N*-terminus is involved in protein function e.g. the *N*-terminal nucleophile hydrolase superfamily where, as the name suggests, the amino group of the *N*-terminal amino acid of the peptide chain acts as a nucleophile in the active site (Mei and Zalkin, 1989). As a result, the enzyme responsible for removal of *N*-terminal methionine, methionine aminopeptidase, is essential for cell viability (Chang et al., 1989). Additionally, methionine aminopeptidase is unable to hydrolyse *N*-formylmethionine (Solbiati et al., 1999), thus deformylase activity is required prior to methionine aminopeptidase activity. For that reason, deformylase activity is also essential for cell viability (Mazel et al., 1994).

The essential nature of deformylation in bacteria has lead to the development of deformylase-inhibitors as antibiotics (Clements et al., 2001; Giglione et al., 2000). Natural deformylase-inhibitors are produced by various organisms: actinonin was identified in 1962 as an antibiotic produced by a novel actinomycete (Gordon et al., 1962), and was later identified as a deformylase-inhibitor (Chen et al., 2000); fumimycin is produced by *Aspergillus fumisynnematus* and shows antibacterial activity against the medically important methicillin and quinolone-resistant *Staphylococcus aureus* (Kwon et al., 2007); two peptide deformylase inhibitors, Sch 382582 and Sch 382583, have been isolated from *Streptomyces sp.* (Chu et al., 2001). The development of deformylase-inhibitors as therapeutic antibiotics was hindered by the discovery of a peptide deformylase homologue in human mitochondria (as well as other

eukaryotic organelles) (Giglione, 2000). Additionally, exposure of human tissue culture cells to actinonin inhibits mitochondrial deformylase activity and prevents cell proliferation (Lee et al., 2004). Despite this, clinical trials are currently being undertaken with the deformylase inhibitor GSK1322322 (O'Dwyer et al., 2013). This anti-proliferative activity of deformylase-inhibitors against human cells is also being investigated for anticancer activity (Lee et al., 2013).

Requirement for formylation

Despite being ubquitous, formylation is not an essential activity in most bacteria. Formylation can be inhibited through the action of the antibiotic trimethoprim. Trimethoprim prevents tetrahydrofolate (THF) production through inhibition of dihydrofolate reductase, consequently preventing production of the formylase cofactor 10-CHO-THF. This inhibition of THF production is usually lethal to bacteria, making trimethoprim an effective antibiotic. However, bacteria can grow in the presence of trimethoprim if supplied with essential metabolites normally synthesised using THF e.g. thymidine, methionine. Such supplementation does not restore THF production, thus formylase activity is still inhibited. Initiator methionyl-$tRNA_i^{fMet}$ produced in *E. coli* cells grown in the presence of trimethoprim (with supplementation of essential metabolites) remains Met-$tRNA_i^{fMet}$. Consequently, translation proceeds with non-formylated Met-$tRNA_i^{fMet}$ - similar to translation in eukaryotes and archaea, albeit with a reduced growth rate (Harvey, 1973).

It is clear that bacteria grown in trimethoprim initiate translation without formylation, however it is difficult to attribute the reduced growth rate of formylation-deficient strains solely to inhibition of formylase activity. Trimethoprim disrupts the synthesis of several important metabolites, and

although these can be supplied in the growth media, the disruption of such biosynthetic pathways may alter the growth of bacteria. Therefore, to understand the effects of translation without formylation, it is important to abolish formylase activity in a more selective manner. Gene knockout experiments have been successful in excising the formylase-encoding gene, *fmt,* from the bacterial chromosome, thus generating truly non-formylating strains (Table 1). As with inhibition using trimethoprim, knocking out the formylase gene decreased the growth rate of all species tested. However, the severity of this growth rate impairment appears species-dependent, with growth rates of non-formylating strains ranging from 0.7x wild-type (for *Enterococcus faecalis* grown in the absence of folate (Samuel et al., 1970)) to complete growth inhibition (the *fmt* gene is essential in *Streptococcus pneumoniae* (Margolis et al., 2001; Song et al., 2005)). This difference in growth rates is likely due to the binding affinity of initiation factors e.g. IF2, for non-formylated Met-tRNA$_i^{fMet}$ (as discussed in section "Formylation" above).

In the absence of formylase activity, the deformylase gene can also be deleted (Mazel et al., 1994). Therefore deformylation is not required when translation is initiated with Met-tRNA$_i^{fMet}$. In contrast, deformylase activity is essential when translation is initiated with fMet-tRNA$_i^{fMet}$ (Mazel et al., 1994).

Table 1 | Growth rate of non-formylating bacteria. The growth rate of *fmt*-knockout (non-formylating) strains generated through different mechanisms reveals consistently lower growth rates than wild-type in all species tested. Growth rates (measured as generation times) of *fmt*- strains are shown relative to generation time of wild-type, formylation proficient strains.

Species	**Growth rate of *fmt*- strain**	**Genotype**	**Reference**
Staphylococcus aureus	0.57	*Δfmt::aac-aph*	Margolis et al., 2000
Haemophilus influenzae	0.50	Deletion of *fmt* base 307	Hackbarth et al., 2002
Bacillus subtilis	0.48	Deletion of *fmt* bases 295-639	Duroc et al., 2009
Pseudomonas aeruginosa	0.36	*fmt::Gm*^{R}	Newton et al., 1999b
Salmonella enterica	0.24	Deletion of *fmt* bases 363-367	Nilsson et al., 2006
Escherichia coli	0.12	*fmtΔI::kan*	Guillon et al., 1992
	~0.2	*Δ(def-fmt):cat*	Mazel et al., 1994
Streptococcus pneumoniae	0*		Margolis et al., 2001; Song et al., 2005

**fmt* gene appears essential in *Streptococcus pneumoniae.*

Proposed functions of formylation

Genes encoding the formylase and deformylase enzymes (*fmt* and *def,* respectively) have been identified in almost all bacterial genomes (Yuan et al., 2001) suggesting that they are ancient genes (Mazel et al., 1997). Formylase and deformylase activity is also found in the bacterial-endosymbiont-derived organelles of eukaryotes – the mitochondria and plastids (Giglione and Meinnel, 2001). Such a degree of conservation suggests that the formylation-deformylation cycle has an important function in translation. Several functions for formylation have been hypothesised and these are discussed further below.

Initiator tRNA identification

It has been hypothesised that formylation of Met-$tRNA_i^{fMet}$ allows discrimination between initiator and elongator tRNAs at the ribosome. Indeed, IF2 has a higher affinity for fMet-$tRNA_i^{fMet}$ than Met-$tRNA_i^{fMet}$ due to electrostatic interactions in the IF2 initiator tRNA binding cleft (see section 'Deformylation' above). However, it appears that structural features of the $tRNA_i^{fMet}$ are more important in identifying the tRNA as a correct initiator molecule. Mutation of any of the three conserved G·C base pairs in the anticodon stem of $tRNA_i^{fMet}$ significantly compromises its ability to function in translation initiation, regardless of formylation (Seong and RajBhandary, 1987a). This is a result of interactions between IF3 and the anticodon stem-loop structure which act to ensure the correct $tRNA_i$ is positioned in the ribosomal P-site (Hartz et al., 1990). Moreover, it appears that *P. aeruginosa* IF2 is able to bind non-formylated Met-$tRNA_i^{fMet}$ almost as efficiently as fMet-$tRNA_i^{fMet}$ (Steiner-Mosonyi et al., 2004). Methionyl-tRNA formylation is probably only a minor component of identification of the initiator tRNA identity.

It is clear that formylation of Met-$tRNA_i^{fMet}$ would prevent the use of initiator tRNA in the elongation phase of translation – the *N*-blocked methionine is not

able to form a peptide bond. However, the disadvantage of using an initiator tRNA in elongation is not obvious; both Met-$tRNA_i^{fMet}$ and Met-$tRNA_e^{Met}$ have the same anticodon and cognate amino acid, so translation should not be disrupted. It has also been shown that prevention of formylation by incorporation of a $C_1 \cdot G_{72}$ in $tRNA_i^{fMet}$ is sufficient to allow use of Met-$tRNA_i^{fMet}$ in elongation (Seong and RajBhandary, 1987b). Additionally, mammalian mitochondria encode a single Met-tRNA species which is only partially formylated. The formylated Met-tRNA molecules are used in translation initiation, and the non-formylated Met-tRNA molecules are used in elongation (Spencer and Spremulli, 2004; Takemoto et al., 1995).

Finally, eukaryotes and archaea do not formylate Met-$tRNA_i^{Met}$, indicating that formylation is not a necessary feature for the differentiation of initiator and elongator Met-tRNA.

Prevention of side-reactions

A second hypothesis for the function of formylation is that it prevents unfavourable side reactions involving the reactive amino group of methionine (Bingel-Erlenmeyer et al., 2008). Indeed, formylation of Met-$tRNA_i^{fMet}$ does reduce the sensitivity of the aminoacyl-linkage to hydrolysis (Schofield and Zamecnik, 1968). However, any unfavourable side reactions that may compromise Met-$tRNA_i^{fMet}$ would also compromise the non-formylated elongator Met-$tRNA_e^{Met}$. There is no reason to suggest that the charged initiator tRNA requires protection to a greater extent than the charged elongator tRNA.

A further site where side-reactions may compromise the integrity of methionine would be the nascent polypeptide. If translation in bacteria proceeded with a non-formylated Met-$tRNA_i^{fMet}$, the nascent polypeptide produced during translation would also have a non-formylated *N*-terminus. However, the extent of protection provided by formylation at this site would be shortlived; the peptide deformylase enzyme binds to the ribosomal exit site, deformylating the nascent peptide as it

emerges from the ribosome. Deformylase activity reaches completion before 48 amino acids of the nascent peptide have been synthesised (Sandikci et al., 2013), suggesting that any protective benefit of formylation would have to be required within the ribosomal exit tunnel. Such a scenario seems unlikely, especially when one considers the amino acids arginine and lysine both have side chains terminating in amino-groups. There is no requirement for protection of these reactive amino groups, hence it follows that there should be no requirement for protection of the *N*-terminal amino group of nascent peptides.

Moreover, the Met-tRNA$_i^{Met}$ of eukaryotes and archaea shows no requirement of formylation for protection from side reactions.

Polar translation of polycistronic mRNA

A third hypothesis for the function of formylation considers the expression of polycistronic mRNA in bacteria. It is proposed that the difference in utilisation of formylation between bacteria and eukaryotes mirrors the use of polycistronic mRNA in these domains. That is, bacteria utilise formylation and transcribe many genes in polycistronic messages, whereas eukaryotes do not use formylation and almost always use monocistronic mRNAs (Kozak, 1999). The difference is explained by the observation that purified bacterial 30S ribosomes have a similar binding affinity for both non-formylated Met-tRNA$_i^{fMet}$ and fMet-tRNA$_i^{fMet}$, whereas 70S ribosomes have a strong affinity bias towards fMet-tRNA$_i^{fMet}$ (Petersen et al., 1976b). The binding of fMet-tRNA$_i^{fMet}$ to 70S ribosomes was also observed to shift an equilibrium of ribosome conformations towards a more active state (Petersen et al., 1976a). These results lead to the hypothesis that translation initiation using dissociated ribosomes can proceed using non-formylated Met-tRNA$_i^{fMet}$, whereas translation initiation using 70S ribosomes is restricted to fMet-tRNA$_i^{fMet}$ (Petersen et al., 1976b). Additionally, it was shown that translation of the *E. coli lac* operon was more coordinated (upstream genes translated before downstream genes) in the presence of formylation than in the

absence (Petersen et al., 1978). Indeed, the formylation state of bacteria may alter ribosome assembly dynamics; IF2 is responsible for binding the formylated methionine of fMet-$tRNA_i^{fMet}$, and interacts with both the 30S and 50S ribosomal subunits during assembly (Antoun et al., 2003). However, a major limitation in interpreting the evidence leading to these results is that many of the assays were performed using purified components of the translation machinery. It has since been shown that the presence of initiation factors greatly alters the affinity of the initiation complex for further factors, including the initiator tRNA (Milón et al., 2012). Additionally, IF2 associates with the 30S ribosomal subunit at a much faster rate than $tRNA_i^{fMet}$, suggesting that non-formylated Met-$tRNA_i^{fMet}$ would have little chance to interact with the ribosome before the formylation-discriminating IF2 was bound (Milón et al., 2012). As such, the results observed with dissociated ribosomes may not apply *in vivo*. Furthermore, in wild type *E. coli,* translation reinitiation on polycistronic operons appears to proceed through an IF3 dependent pathway, suggesting the use of dissociated ribosomes rather than 70S ribosomes as this hypothesis predicts (Yoo and RajBhandary, 2008).

Finally, polycistronic messages are common in archaea (Wurtzel et al., 2010) and show polar expression (Santangelo et al., 2008) without a reliance on formylation.

All of the currently proposed theories on the function of Met-$tRNA_i$ formylation have been developed from an 'adaptionist' viewpoint. That is, the presence of formylation has been explained by attempting to attribute a fitness advantage to cells utilizing the process. These hypothesised advantages of formylation all suffer the same shortcoming – the formylation-deformylation cycle is not required for efficient translation in either eukaryotes or archaea, thus it is difficult to argue it is an essential feature of translation initiation. Hence, any reasonable fitness

benefit conferred by formylation (or indeed, any fitness deficit caused by a lack of formylation) must be specific to bacteria.

Evolution of non-adaptive traits

As we have seen with formylation, it has been tempting to reduce traits down to a singular process and attribute their presence to an adaptive benefit the trait confers upon the organism e.g. formylation preventing unfavourable side reactions of methionine. Such 'adaptionist' reasoning assumes traits evolve by purely adaptive processes that tend to produce the optimally adapted organism. This reasoning was criticised by Gould and Lewontin as approaching a 'Panglossian paradigm' (Gould and Lewontin, 1979). In Voltaire's novel Candide, Dr. Pangloss teaches:

> "It is demonstrable... that things cannot be otherwise than as they are; for all being created for an end, all is necessarily for the best end. Observe, that the nose has been formed to bear spectacles—thus we have spectacles. Legs are visibly designed for stockings—and we have stockings."

Gould and Lewontin suggested that this reasoning mirrors, to some degree, the ideologies used in attempting to understand biological systems. That is, rather than speculating on the evolutionary origin of traits, we tend to concentrate on small functional details in an attempt to reverse engineer complex biological systems. This method of reasoning may be suitable when the origin of the phenotype is known and confers a clear adaptive advantage, such is the case with industrial melanism in the peppered moth, *Biston betularia*. Here, a spontaneous mutation in genes governing colour-patterning produces a darker wing phenotype (van't Hof et al., 2011) consequently decreasing predation in geographical regions with high levels of air pollution (Lees and Creed, 1975). In this case, the adaptive role for the trait is clear: the genotype encoding melanic wing forms provides a selective advantage by reducing predation, thus the genotype increases in frequency. However, when the selective advantage conferred by a trait is not clear, this adaptation-based reasoning can limit potential hypotheses for the

origin of a phenotype. An example of this can be seen in bacterial toxin-antitoxin (TA) loci. Here, a gene pair encoding a stable toxin and labile antitoxin is able to persist in cells without conferring an obvious functional benefit. If expression of the TA gene pair is disrupted, the antitoxin will break down faster than the toxin, and the cell will succumb to the action of the toxin (Van Melderen and De Bast, 2009). Through this process, cells can become 'addicted' to TA expression. The numerous chromosomal TA loci in bacteria are often regarded as stress-response loci (due to their activation under stress conditions), responsible for 'programmed cell death' in response to nutrient limitation (Rice and Bayles, 2003). The altruistic death of individuals is thought to prevent the death of entire populations under nutrient stress, thus conferring a selective advantage upon hosts. Despite this prevailing theory, TA loci have been shown to confer a cost upon host cells (Cooper and Heinemann, 2005; Mochizuki et al., 2006), and no fitness advantage has been observed for strains carrying TA loci upon exposure to stress (Tsilibaris et al., 2007). Relying on this cell-adaptation view excludes pathways of evolution that may help explain the spread of TA pairs without requiring a benefit to the host e.g. plasmid-encoded TA loci facilitate competition between horizontally mobile elements (Cooper and Heinemann, 2000; 2005; Cooper et al., 2010; Heinemann, 1998); TA loci can prematurely abort phage infection (Fineran et al., 2009; Pecota and Wood, 1996); TA loci stabilize otherwise dispensable integrative elements (Szekeres et al., 2007; Wozniak and Waldor, 2009). Under these models, the presence of TA loci is not explained by their ability to increase the fitness of hosts, but rather their pervasive horizontal transfer and subsequent addictive phenotype. In attempting to determine the evolutionary pathway by which modern traits have emerged, we must be careful not to limit our view to one of in which the trait serves the cell.

Dr. Pangloss's quote also draws on another difficulty in evolutionary biology – differentiation between current utility and evolutionary origin. It is obvious that

Dr. Pangloss is mistaken; the nose did not form to bear spectacles, nor the legs to wear stockings, however in evolutionary biology these distinctions can be less clear. The evolutionary forces by which traits arise may not be reflected in their function in modern cells. In some cases, the difference between current utility and evolutionary origin is clear; such as the case with the *RAG1* gene involved in the vertebrate immune system. This gene is essential to the process of V(D)J recombination in the production of antibodies, and thus, adaptive immunity (Litman et al., 2010). However, the origin of *RAG1* is not related to vertebrate immunity, rather the gene was likely derived from a transposase responsible for movement of the *Transib* transposon (Kapitonov and Jurka, 2005).

In many cases, however, the differentiation between current utility and evolutionary origin is more difficult, as is the case with TA loci. It is not implausible that TA loci function in stress-responsive programmed cell death, however, if this is indeed a current function of chromosomal TA loci, we must ask how such a system could evolve and whether stress response is what conditioned its pathway of evolution. It seems unlikely that the selective advantage realised in sporadic times of stress would outweigh the continual cost of encoding a TA loci, unless stress conditions were a common occurrence (in which case we might expect the appearance of traits to alleviate stress). Additionally, any individuals that lose the TA loci would reap the benefit of their altruistic TA encoding neighbours, without risking death. Almost counter-intuitively, in the presence of 'programmed cell death', it is advantageous not to carry the trait (Köhler et al., 2009; Velicer et al., 2000). If indeed TA loci function in altruistic stress response, it seems likely that this is an exaptation of systems that have evolved through another mechanism such as those mentioned above.

The fact that formylation is not necessary for translation in any of the three domains of life (it is absent from eukaryotes and archaea and dispensible from most bacteria) suggests it is not a critical feature of translation initiation.

Additionally, the identification of features which allow translation in the absence of formylation (Steiner-Mosonyi et al., 2004) suggest the fitness disadvantage observed upon loss of formylation may be at least partially overcome. Moreover, formylation is ubiquitous in the bacterial domain, yet completely absent from archaeal and eukaryotic cytoplasmic translation. Despite this, there is no obvious reason why it would not function in eukaryotes and archaea. In light of extensive interdomain horizontal gene transfer, this strict distribution of formylation is curious, and may suggest an evolutionary mechanism linked to bacteria rather than to an advantageous function. Consequently, adaptive hypotheses for the function of formylation are unlikely to fully explain its evolution.

Summary

It is clear that the function of Met-tRNA$_i$ formylation in bacterial translation is not understood. Investigations into specific roles that formylation may have within the cell can provide insight into the current utility of formylation, however they provide little information regarding the evolution of the system. The absence of formylation from two domains of life makes it difficult to argue for an intrinsic requirement for fMet-tRNA$_i$ in translation. Perhaps in looking for a functional role of formylation, previous studies have looked past a non-adaptive evolutionary history by which the formylation-deformylation system could have evolved.

The primary aim of this study was to understand how the formylation-deformylation cycle evolved in bacteria. To answer this, we asked whether formylation was required for efficient translation in bacteria, or whether, in a manner similar to eukaryotes and archaea, bacteria could synthesise proteins effectively in the absence of formylation. Using a long-term culturing technique we isolated non-formylating *E. coli* with growth rates indistinguishable from wild-type, formylating cells. This enabled us to examine the genes involved in translation and how they are affected by the absence of formylation. Following this, we examined the effect of introducing the formylase-deformylase gene pair into non-formylating, growth rate compensated bacteria, and formylation-naïve yeast cells on a plasmid vector. Finally, we investigated whether genes encoding formylation and deformylation activity could be found in archaea.

References

Adams, J.M. (1968). On the release of the formyl group from nascent protein. J. Mol. Biol. *33*, 571–589.

Andreou, A.Z., and Klostermeier, D. (2014). eIF4B and eIF4G Jointly Stimulate eIF4A ATPase and Unwinding Activities by Modulation of the eIF4A Conformational Cycle. J. Mol. Biol. *426*, 51–61.

Antoun, A., Pavlov, M.Y., Andersson, K., Tenson, T., and Ehrenberg, M. (2003). The roles of initiation factor 2 and guanosine triphosphate in initiation of protein synthesis. EMBO J. *22*, 5593–5601.

Battiste, J.L., Pestova, T.V., Hellen, C.U., and Wagner, G. (2000). The eIF1A solution structure reveals a large RNA-binding surface important for scanning function. Mol. Cell *5*, 109–119.

Benelli, D., Maone, E., and Londei, P. (2003). Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. Mol. Microbiol. *50*, 635–643.

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolić, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B., and Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature *452*, 108–111.

Chang, S.Y., McGary, E.C., and Chang, S. (1989). Methionine aminopeptidase gene of Escherichia coli is essential for cell growth. J. Bacteriol. *171*, 4071–4072.

Chen, D.Z., Patel, D.V., Hackbarth, C.J., Wang, W., Dreyer, G., Young, D.C., Margolis, P.S., Wu, C., Ni, Z.-J., Trias, J., et al. (2000). Actinonin, a Naturally Occurring Antibacterial Agent, Is a Potent Deformylase Inhibitor. Biochemistry *39*, 1256–1262.

Chu, M., Mierzwa, R., He, L., Xu, L., Gentile, F., Terracciano, J., Patel, M., Miesel, L., Bohanon, S., Kravec, C., et al. (2001). Isolation and structure elucidation of two novel deformylase inhibitors produced by Streptomyces sp. Tetrahedron Letters *42*, 3549–3551.

Clements, J.M., Beckett, R.P., Brown, A., Catlin, G., Lobell, M., Palan, S., Thomas, W., Whittaker, M., Wood, S., Salama, S., et al. (2001). Antibiotic activity and characterization of BB-3497, a novel peptide deformylase inhibitor. Antimicrob. Agents Chemother. *45*, 563–570.

Coleman, J. (1992). Characterization of the Escherichia coli gene for 1-acyl-sn-glycerol-3-phosphate acyltransferase (pIsC). Molec. Gen. Genet. *232*, 295–303.

Cooper, T.F., and Heinemann, J.A. (2000). Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. PNAS *97*, 12643–12648.

Cooper, T.F., and Heinemann, J.A. (2005). Selection for plasmid post-segregational killing depends on multiple infection: evidence for the selection of more virulent parasites through parasite-level competition. Proc. R. Soc. B *272*, 403–410.

Cooper, T.F., Paixão, T., and Heinemann, J.A. (2010). Within-host competition selects for plasmid-encoded toxin-antitoxin systems. Proc. R. Soc. B *277*, 3149–3155.

Dallas, A., and Noller, H.F. (2001). Interaction of Translation Initiation Factor 3 with the 30S Ribosomal Subunit. Mol. Cell *8*, 855–864.

Dennis, P.P. (1997). Ancient ciphers: translation in Archaea. Cell 89, 1007-1010.

Duroc, Y., Giglione, C., and Meinnel, T. (2009). Mutations in three distinct loci cause resistance to peptide deformylase inhibitors in Bacillus subtilis. Antimicrob. Agents Chemother. *53*, 1673–1678.

Fabbretti, A., Pon, C.L., Hennelly, S.P., Hill, W.E., Lodmell, J.S., and Gualerzi, C.O. (2007). The real-time path of translation factor IF3 onto and off the ribosome. Mol. Cell *25*, 285–296.

Filbin, M.E., and Kieft, J.S. (2009). Toward a structural understanding of IRES RNA function. Curr. Opin. Struct. Biol. *19*, 267–276.

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., and Salmond, G.P.C. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. PNAS *106*, 894–899.

Fox, G.E., and Naik, A.K. (2004). The Evolutionary History of the Translation Machinery. In The Genetic Code and the Origin of Life, (Boston, MA: Springer US), pp. 92–105.

Gabashvili, I.S., Agrawal, R.K., Spahn, C.M., Grassucci, R.A., Svergun, D.I., Frank, J., and Penczek, P. (2000). Solution structure of the E. coli 70S ribosome at 11.5 A resolution. Cell *100*, 537–549.

Garofalo, C., Kramer, G., and Appling, D.R. (2005). Characterization of the C2

subdomain of yeast mitochondrial initiation factor 2. Arch. Biochem. Biophys. *439*, 113–120.

Giglione, C. (2000). Identification of eukaryotic peptide deformylases reveals universality of N-terminal protein processing mechanisms. EMBO J. *19*, 5916–5929.

Giglione, C., and Meinnel, T. (2001). Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. Trends Plant Sci. *12(6)*, 566-572

Giglione, C., Pierre, M., and Meinnel, T. (2000). Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. Mol. Microbiol. *36*, 1197–1205.

Gordon, J.J., Kelly, B.K., and Miller, G.A. (1962). Actinonin: an antibiotic substance produced by an actinomycete. Nature *195*, 701–702.

Gould, P.S., Dyer, N.P., Croft, W., Ott, S., and Easton, A.J. (2014). Cellular mRNAs access second ORFs using a novel amino acid sequence-dependent coupled translation termination-reinitiation mechanism. RNA *20*, 373–381.

Gould, S.J., and Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R. Soc. Lond., B, Biol. Sci. *205*, 581–598.

Guan, Z., Wang, X., and Raetz, C.R.H. (2011). Identification of a chloroform-soluble membrane miniprotein in Escherichia coli and its homolog in Salmonella typhimurium. Anal. Biochem. *409*, 284–289.

Guenneugues, M., Caserta, E., Brandi, L., Spurio, R., Meunier, S., Pon, C.L., Boelens, R., and Gualerzi, C.O. (2000). Mapping the fMet-tRNAfMet binding site of initiation factor IF2. EMBO J. *19*, 5233–5240.

Guillon, J.M., Mechulam, Y., Schmitter, J.M., Blanquet, S., and Fayat, G. (1992). Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of Escherichia coli. J. Bacteriol. *174*, 4294–4301.

Hackbarth, C.J., Chen, D.Z., Lewis, J.G., Clark, K., Mangold, J.B., Cramer, J.A., Margolis, P.S., Wang, W., Koehn, J., Wu, C., et al. (2002). N-alkyl urea hydroxamic acids as a new class of peptide deformylase inhibitors with antibacterial activity. Antimicrob. Agents Chemother. *46*, 2752–2764.

Hartz, D., Binkley, J., Hollingsworth, T., and Gold, L. (1990). Domains of initiator tRNA and initiation codon crucial for initiator tRNA selection by

Escherichia coli IF3. Genes Dev. *4*, 1790–1800.

Harvey, R.J. (1973). Growth and Initiation of Protein Synthesis in Escherichia coli in the Presence of Trimethoprim. J. Bacteriol. *114,* 309-322.

Hasenöhrl, D., Benelli, D., BARBAZZA, A., Londei, P., and Bläsi, U. (2006). Sulfolobus solfataricus translation initiation factor 1 stimulates translation initiation complex formation. RNA *12*, 674–682.

Hasenöhrl, D., Fabbretti, A., Londei, P., Gualerzi, C.O., and Bläsi, U. (2009). Translation initiation complex formation in the crenarchaeon Sulfolobus solfataricus. RNA *15*, 2288–2298.

Heinemann, J.A. (1998). Looking sideways at the evolution of replicons. In Horizontal Gene Transfer, M. Syvanen, and C.I. Kado, eds. (London: Academic Press), pp. 11–24.

Hinnebusch, A.G. (2006). eIF3: a versatile scaffold for translation initiation complexes. Trends Biochem. Sci. *31*, 553–562.

Hoeppner, M.P., Gardner, P.P., and Poole, A.M. (2012). Comparative analysis of RNA families reveals distinct repertoires for each domain of life. PLoS Comput. Biol. *8*, e1002752.

Hsiao, C., Mohan, S., Kalahar, B.K., and Williams, L.D. (2009). Peeling the onion: ribosomes are ancient molecular fossils. Mol. Biol. Evol. *26*, 2415–2425.

Kapitonov, V.V., and Jurka, J. (2005). RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. PLoS Biol. *3*, e181.

Karamyshev, A.L., Karamysheva, Z.N., Yamami, T., Ito, K., and Nakamura, Y. (2004). Transient idling of posttermination ribosomes ready to reinitiate protein synthesis. Biochimie *86*, 933–938.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. Gene *234*, 187–208.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. *15*, 8125–8148.

Köhler, T., Buckling, A., and van Delden, C. (2009). Cooperation and virulence of clinical Pseudomonas aeruginosa populations. PNAS *106*, 6339–6344.

Kramer, P., Gäbel, K., Pfeiffer, F., and Soppa, J. (2014). Haloferax volcanii, a Prokaryotic Species that Does Not Use the Shine Dalgarno Mechanism for

Translation Initiation at 5′-UTRs. PLoS ONE *9*, e94979.

Kwon, Y.-J., Sohn, M.-J., Zheng, C.-J., and Kim, W.-G. (2007). Fumimycin: a peptide deformylase inhibitor with an unusual skeleton produced by Aspergillus fumisynnematus. Org. Lett. *9*, 2449–2451.

Kyrpides, N.C., and Woese, C.R. (1998). Universally conserved translation initiation factors. PNAS *95*, 224–228.

Lee, C.P., Seong, B.L., and RajBhandary, U.L. (1991). Structural and sequence elements important for recognition of Escherichia coli formylmethionine tRNA by methionyl-tRNA transformylase are clustered in the acceptor stem. J. Biol. Chem. *266*, 18012–18017.

Lee, J.H., Pestova, T.V., Shin, B.-S., Cao, C., Choi, S.K., and Dever, T.E. (2002). Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. PNAS *99*, 16689–16694.

Lee, M.D., She, Y., Soskis, M.J., Borella, C.P., Gardner, J.R., Hayes, P.A., Dy, B.M., Heaney, M.L., Philips, M.R., Bornmann, W.G., et al. (2004). Human mitochondrial peptide deformylase, a new anticancer target of actinonin-based antibiotics. J. Clin. Invest. *114*, 1107–1116.

Lee, S.J., Jung, O.S., Lee, B.J., Cho, K.H., and Lee, B.I. (2013). Identification of Potent Inhibitors against Human Peptide Deformylase as Anticancer Agents. Bull. Korean Chem. Soc. *34(12)* 3885-3887.

Lees, D.R., and Creed, E.R. (1975). Industrial Melanism in Biston betularia: The Role of Selective Predation. J. Anim. Ecol. *44*, 67.

Li, L., and Wang, C.C. (2004). Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, Giardia lamblia. J. Biol. Chem. *279*, 14656–14664.

Litman, G.W., Rast, J.P., and Fugmann, S.D. (2010). The origins of vertebrate adaptive immunity. Nat. Rev. Immunol. *10*, 543–553.

Marcker, K., and Sanger, F. (1964). N-Formyl-methionyl-S-RNA. J. Mol. Biol. *8*, 835–838.

Margolis, P.S., Hackbarth, C.J., Young, D.C., Wang, W., Chen, D., Yuan, Z., White, R., and Trias, J. (2000). Peptide deformylase in Staphylococcus aureus: resistance to inhibition is mediated by mutations in the formyltransferase gene. Antimicrob. Agents Chemother. *44*, 1825–1831.

Margolis, P., Hackbarth, C., Lopez, S., Maniar, M., Wang, W., Yuan, Z., White, R., and Trias, J. (2001). Resistance of Streptococcus pneumoniae to deformylase inhibitors is due to mutations in defB. Antimicrob. Agents Chemother. *45*, 2432–2435.

Matheson, A.T., and Yaguchi, M. (1975). The conservation of amino acids in the N-terminal position of ribosomal and cytosol proteins from Escherichia coli, Bacillus stearothermophilus, and Halobacterium cutirubrum. Can. J. Biochem. *53* 1323-1327.

Mazel, D., Coïc, E., Blanchard, S., and Saurin, W. (1997). A survey of polypeptide deformylase function throughout the eubacterial lineage. J. Mol. Biol. *266.* 939-949.

Mazel, D., Pochet, S., and Marliere, P. (1994). Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO J. *13*, 914–923.

Mei, B., and Zalkin, H. (1989). A cysteine-histidine-aspartate catalytic triad is involved in glutamine amide transfer function in purF-type glutamine amidotransferases. J. Biol. Chem. *264*, 16613–16619.

Meinnel, T., and Blanquet, S. (1995). Enzymatic properties of Escherichia coli peptide deformylase. J. Bacteriol. *177*, 1883–1887.

Meunier, S., Spurio, R., Czisch, M., Wechselberger, R., Guenneugues, M., Gualerzi, C.O., and Boelens, R. (2000). Structure of the fMet-tRNAfMet-binding domain of B.stearothermophilus initiation factor IF2. EMBO J. *19*, 1918–1926.

Milligan, D.L., and Koshland, D.E. (1990). The amino terminus of the aspartate chemoreceptor is formylmethionine. J. Biol. Chem. *265,* 4455-4460.

Milón, P., and Rodnina, M.V. (2012). Kinetic control of translation initiation in bacteria. Crit. Rev. Biochem. Mol. Biol. *47*, 334–348.

Milón, P., Carotti, M., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., and Gualerzi, C.O. (2010). The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. EMBO Reports *11*, 312–316.

Milón, P., Maracci, C., Filonava, L., Gualerzi, C.O., and Rodnina, M.V. (2012). Real-time assembly landscape of bacterial 30S translation initiation complex. Nat. Struct. Mol. Biol. *19*, 609–615.

Mochizuki, A., Yahara, K., Kobayashi, I., and Iwasa, Y. (2006). Genetic addiction: selfish gene's strategy for symbiosis in the genome. Genetics *172*, 1309–

1323.

Moll, I., Hirokawa, G., Kiel, M.C., Kaji, A., and Bläsi, U. (2004). Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. Nucleic Acids Res. *32*, 3354–3363.

Nakamura, Y., Ito, K., and Isaksson, L.A. (1996). Emerging understanding of translation termination. Cell *87*, 147–150.

Newton, D.T., Creuzenet, C., and Mangroo, D. (1999a). Formylation is not essential for initiation of protein synthesis in all eubacteria. J. Biol. Chem. *274*, 22143–22416.

Newton, D.T., Creuzenet, C., and Mangroo, D. (1999b). Formylation is not essential for initiation of protein synthesis in all eubacteria. J. Biol. Chem. *274*, 22143–22146.

Newton, D.T., Niemkiewicz, M., Lo, R.Y.C., and Mangroo, D. (1999c). Recognition of the initiator tRNA by the Pseudomonas aeruginosa methionyl-tRNA formyltransferase: importance of the base-base mismatch at the end of the acceptor stem. FEMS Microbiol. Lett. *178*, 289–298.

Nilsson, A.I., Zorzet, A., Kanth, A., Dahlström, S., Berg, O.G., and Andersson, D.I. (2006). Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. PNAS *103*, 6976–6981.

O'Dwyer, K., Hackel, M., Hightower, S., Hoban, D., Bouchillon, S., Qin, D., Aubart, K., Zalacain, M., and Butler, D. (2013). Comparative analysis of the antibacterial activity of a novel peptide deformylase inhibitor, GSK1322322. Antimicrob. Agents Chemother. *57*, 2333–2342.

Oberer, M., Marintchev, A., and Wagner, G. (2005). Structural basis for the enhancement of eIF4A helicase activity by eIF4G. Genes Dev. *19*, 2212–2223.

Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R., and Ramakrishnan, V. (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. Mol. Cell *26*, 41–50.

Pecota, D.C., and Wood, T.K. (1996). Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. J. Bacteriol. *178*, 2044–2050.

Petersen, H.U., Danchin, A., and Grunberg-Manago, M. (1976a). Toward an understanding of the formylation of initiator tRNA methionine in prokaryotic protein synthesis. II. A two-state model for the 70S ribosome. Biochem. *15(7),*

1362-1369.

Petersen, H.U., Joseph, E., Ullmann, A., and Danchin, A. (1978). Formylation of initiator tRNA methionine in procaryotic protein synthesis: in vivo polarity in lactose operon expression. J. Bacteriol. *135*, 453–459.

Petersen, H.U., Danchin, A., and Grunberg-Manago, M. (1976b). Toward an understanding of the formylation of initiator tRNA methionine in prokaryotic protein synthesis. I. In vitro studies of the 30S and 70S ribosomal-tRNA complex. Biochem. *15*, 1357–1362.

Petrov, A.S., Bernier, C.R., Hsiao, C., Norris, A.M., Kovacs, N.A., Waterbury, C.C., Stepanov, V.G., Harvey, S.C., Fox, G.E., Wartell, R.M., et al. (2014). Evolution of the ribosome at atomic resolution. PNAS *111*, 10251–10256.

Pon, C.L., and Gualerzi, C.O. (1986). Mechanism of translational initiation in prokaryotes. FEBS Letters *195*, 215–219.

Qin, H., Grigoriadou, C., and Cooperman, B.S. (2009). Interaction of IF2 with the Ribosomal GTPase-Associated Center during 70S Initiation Complex Formation. Biochem. *48*, 4699–4706.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. Cell *108*, 557–572.

Rice, K.C., and Bayles, K.W. (2003). Death's toolbox: examining the molecular components of bacterial programmed cell death. Mol. Microbiol. *50*, 729–738.

Rogers, G.W., Richter, N.J., Lima, W.F., and Merrick, W.C. (2001). Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. J. Biol. Chem. *276*, 30914–30922.

Samuel, C.E., D'Ari, L., and Rabinowitz, J.C. (1970). Evidence against the folate-mediated formylation of formyl-accepting methionyl transfer ribonucleic acid in Streptococcus faecalis R. J. Biol. Chem. *245*, 5115–5121.

Sandikci, A., Gloge, F., Martinez, M., Mayer, M.P., Wade, R., Bukau, B., and Kramer, G. (2013). Dynamic enzyme docking to the ribosome coordinates N-terminal processing with polypeptide folding. Nat Struct Mol Biol *20*, 843–850.

Santangelo, T.J., Cubonová, L., Matsumi, R., Atomi, H., Imanaka, T., and Reeve, J.N. (2008). Polarity in archaeal operon transcription in Thermococcus kodakaraensis. J. Bacteriol. *190*, 2244–2248.

Sartorius Neef, S., and Pfeifer, F. (2004). In vivo studies on putative Shine–

Dalgarno sequences of the halophilic archaeon Halobacterium salinarum. Mol. Microbiol. *51*, 579–588.

Schmeing, T.M., and Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. Nature *461*, 1234–1242.

Schmitt, E., Panvert, M., Blanquet, S., and Mechulam, Y. (1998). Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. EMBO J. *17*, 6819–6826.

Schofield, P., and Zamecnik, P.C. (1968). Cupric ion catalysis in hydrolysis of aminoacyl-tRNA. Biochim. Biophys. Acta *155*, 410–416.

Schomburg, D., and Stephan, D. (1996). Methionyl-tRNA formyltransferase. In Springer Handbook of Enzymes, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 511–515.

Seong, B.L., and RajBhandary, U.L. (1987a). Escherichia coli formylmethionine tRNA: mutations in GGGCCC sequence conserved in anticodon stem of initiator tRNAs affect initiation of protein synthesis and conformation of anticodon loop. PNAS *84*, 334–338.

Seong, B.L., and RajBhandary, U.L. (1987b). Mutants of Escherichia coli formylmethionine tRNA: a single base change enables initiator tRNA to act as an elongator in vitro. PNAS *84*, 8859–8863.

Shuman, S. (2002). What messenger RNA capping tells us about eukaryotic evolution. Nat Rev Mol Cell Biol *3*, 619–625.

Smith, A.E., and Marcker, K.A. (1970). Cytoplasmic methionine transfer RNAs from eukaryotes. Nature *226*, 607–610.

Solbiati, J., Chapman-Smith, A., Miller, J.L., Miller, C.G., and Cronan, J.E., Jr (1999). Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal. J. Mol. Biol. *290*, 607–614.

Song, J.-H., Ko, K.S., Lee, J.-Y., Baek, J.Y., Oh, W.S., Yoon, H.S., Jeong, J.-Y., and Chun, J. (2005). Identification of essential genes in Streptococcus pneumoniae by allelic replacement mutagenesis. Mol. Cells *19*, 365–374.

Spencer, A.C., and Spremulli, L.L. (2004). Interaction of mitochondrial initiation factor 2 with mitochondrial fMet-tRNA. Nucleic Acids Res. *32*, 5464–5470.

Spurio, R., Brandi, L., Caserta, E., Pon, C.L., Gualerzi, C.O., Misselwitz, R., Krafft, C., Welfle, K., and Welfle, H. (2000). The C-terminal Subdomain (IF2 C-

2) Contains the Entire fMet-tRNA Binding Site of Initiation Factor IF2. J. Biol. Chem. *275*, 2447–2454.

Steiner-Mosonyi, M., Creuzenet, C., Keates, R.A.B., Strub, B.R., and Mangroo, D. (2004). The Pseudomonas aeruginosa initiation factor IF-2 is responsible for formylation-independent protein initiation in P. aeruginosa. J. Biol. Chem. *279*, 52262–52269.

Studer, S.M., and Joseph, S. (2006). Unfolding of mRNA Secondary Structure by the Bacterial Translation Initiation Complex. Mol. Cell *22*, 105–115.

Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe Magnus, D.A. (2007). Chromosomal toxin–antitoxin loci can diminish large-scale genome reductions in the absence of selection. Mol. Microbiol. *63*, 1588–1605.

Szkaradkiewicz, K., Zuleeg, T., Limmer, S., and Sprinzl, M. (2000). Interaction of fMet-tRNAfMet and fMet-AMP with the C-terminal domain of Thermus thermophilus translation initiation factor 2. Eur. J. Biochem. *267*, 4290–4299.

Takemoto, C., Koike, T., Yokogawa, T., Benkowski, L., Spremulli, L.L., Ueda, T.A., Nishikawa, K., and Watanabe, K. (1995). The ability of bovine mitochondrial transfer RNAMet to decode AUG and AUA codons. Biochimie *77*, 104–108.

Tsilibaris, V., Maenhaut-Michel, G., Mine, N., and Van Melderen, L. (2007). What is the benefit to Escherichia coli of having multiple toxin-antitoxin systems in its genome? J. Bacteriol. *189*, 6101–6108.

Van Melderen, L., and De Bast, M.S. (2009). Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities? PLoS Genet *5*, e1000437.

van't Hof, A.E., Edmonds, N., Dalíková, M., Marec, F., and Saccheri, I.J. (2011). Industrial melanism in British peppered moths has a singular and recent mutational origin. Science *332*, 958–960.

Velicer, G.J., Kroos, L., and Lenski, R.E. (2000). Developmental cheating in the social bacterium : Myxococcus xanthus. Nature *404*, 598–601.

Weisser, M., Voigts-Hoffmann, F., Rabl, J., Leibundgut, M., and Ban, N. (2013). The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. Nat Struct Mol Biol *20*, 1015–1017.

White, B.N., and Bayley, S.T. (1972). Methionine transfer RNAs from the extreme halophile, Halobacterium cutirubrum. Biochimica Et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis *272*, 583–587.

Wozniak, R.A.F., and Waldor, M.K. (2009). A Toxin–Antitoxin System Promotes the Maintenance of an Integrative Conjugative Element. PLoS Genet *5*, e1000439.

Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B.A., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. Genome Res. *20*, 133–141.

Yoo, J.-H., and RajBhandary, U.L. (2008). Requirements for translation re-initiation in Escherichia coli: roles of initiator tRNA and initiation factors IF2 and IF3. Mol. Microbiol. *67*, 1012–1026.

Yuan, Z., Trias, J., and White, R.J. (2001). Deformylase as a novel antibacterial target. Drug Discovery Today *6*, 954–961.

Chapter 2
The *def-fmt* operon is dispensable from *E. coli*

All experimental and bioinformatic analyses outlined in this chapter were performed by Ryan Catchpole under the mentorship of Prof. Jack Heinemann and Assoc. Prof. Anthony Poole.

Introduction

Bacteria initiate translation with a formylated methionyl tRNA (fMet-$tRNA_i^{fMet}$) (Marcker and Sanger, 1964). In eukaryotic and archaeal translation, the initiating species is the non-formylated Met-$tRNA_i^{Met}$ (Smith and Marcker, 1970; White and Bayley, 1972). Although several functions for the formylation of Met-$tRNA_i^{fMet}$ have been proposed, the absence of formylation from eukaryotic and archaeal translation makes it difficult to argue that it is an essential feature of translation initiation.

Wild-type strains of *E. coli* add the formyl group to Met-$tRNA_i^{fMet}$ using the enzyme methionyl-tRNA formyltransferase (formylase). The resulting fMet-$tRNA_i^{fMet}$ is recognized by initiation factor 2 (IF-2) and used as the initiating tRNA in translation (Guenneugues et al., 2000). Consequently, nascent polypeptides emerge from ribosomes with formylated *N*-termini. Formylated peptides are processed at the ribosomal exit site by peptide deformylase (deformylase), hydrolysing the formyl-group (Meinnel and Blanquet, 1995). The *N*-

terminal methionine residue can then be cleaved by the essential enzyme methionine aminopeptidase (MAP) (Chang et al., 1989). As MAP cannot cleave *N*-blocked methionine residues, deformylation is of fMet-polypeptides is essential (Adams, 1968; Solbiati et al., 1999).

The function of Met-$tRNA_i^{fMet}$ formylation has been described as allowing discrimination between initiator and elongator Met-tRNA, preventing unwanted site reactions involving the amino group of methionine (Bingel-Erlenmeyer et al., 2008), and facilitating polar translation of polycistronic mRNA (Petersen et al., 1976; 1978) (reviewed in chapter 1). However, the restriction of formylation to the bacterial domain suggests any requirement for formylation must be restricted to bacteria, and thus not an inherent requirement in translation itself.

Others had already shown that the *def-fmt* operon could be deleted from *E. coli* (Guillon et al., 1992; Mazel et al., 1994), nevertheless, our examination into the evolution of formylation required a formylation-deficient bacterium with predictable growth characteristics. For this purpose, we chose to delete the *fmt* and *def* genes from the *E. coli* strain REL606. The growth characteristics and genome sequence of this strain has been studied for many decades in laboratory conditions (Wiser et al., 2013) making it an ideal candidate for further evolutionary studies.

Methods

Strains and media

All chemicals were purchased from Sigma-Aldrich Co. unless otherwise specified. All oligonucleotides were synthesized by Integrated DNA Technologies. *E. coli* strain REL606 was obtained from T. Cooper (University of Houston, Texas). REL606 and REL606-derived strains were grown at 37°C in Davis Minimal media (Difco) supplemented with 2mg/L thiamine and either 200mg/L dextrose (DM200) or 2000mg/L dextrose (DM2000), unless otherwise specified. For solid media, bacteriological agar (Oxoid) was added to a final concentration of 1.5%w/v. When required, antibiotics were used at the following concentrations: Streptomycin, 100µg/mL; Kanamycin, 20µg/mL; Ampicillin, 100µg/mL; Actinonin, 100mg/L (Peptides International).

Deleting def-fmt operon

The *def* and *fmt* genes in the REL606 chromosome were deleted by a complete knock out and replaced with an *nptII* cassette using a modified 'Gene Gorging' protocol (Herring et al., 2003). For replacement of the *def-fmt* pair, the *nptII* cassette was amplified from the plasmid pKD4 (Genbank: AY048743.1) (Datsenko and Wanner, 2000) using the following oligonucleotides:

5′ – AGAATAGAAGAAATAATCTTTCTAACTCCTGAACACATCTCTGGAGATTT**GTGTAGGC TGGAGCTGCTTC** – 3′

5′ – *ATTACCCTGTTATCCCTA*TTCATAAGTATAAAAACGCCCGGCAAGACCGGGCTTAGAAG AGTGGACTA**ATGGGAATTAGCCATGGTCC** – 3′

(where bold indicates identity to pKD4, underline indicates identity to *def-fmt* region of *E. coli* chromosome and italics indicates I-*Sce*I recognition sequence).

For replacement of the *fmt* gene, the following oligonucleotides were used:

5′ – GTTGAAAAACTGGATCGTCTGAAAGCCCGGGCTTAAGGATAAGAACTAAC**GTGTAGG CTGGAGCTGCTTC** – 3′

5′ – *ATTACCCTGTTATCCCTA*TTCATAAGTATAAAAACGCCCGGCAAGACCGGGCTTAGAAG AGTGGACTA**ATGGGAATTAGCCATGGTCC** – 3′.

In both cases, the 1614bp product was cloned in to the pGEM-T Easy vector system (Promega) and used as a donor plasmid for gene gorging, with the Lambda recombinase and *I-Sce*I genes supplied by the plasmid pACBSR (Herring et al., 2003). Recombinants were selected on LB agar containing Streptomycin and Kanamycin, then restreaked to DM2000 agar containing Streptomycin and Kanamycin. Colonies showing correct amplification products in PCR screening were confirmed by Sanger sequencing (Macrogen Korea).

Determination of growth rate

Cultures were grown overnight in DM200 (or DM2000 where specified) and diluted 1:100 in fresh media. The OD_{595} of these cultures was then monitored for 16 hours at 37°C (with shaking) using the FLUOstar OPTIMA (BMG Labtech). The growth rate was determined as the minimum doubling time taken over a 30-minute interval.

Complementation

The *def-fmt* operon was amplified by PCR using the following primers:

5′ – CCATAAGCAGCCTTAGCAATCTTTGC – 3′ and

5′ – CAAGACCGGGCTTAGAAGAGTGG – 3′

The resulting 1615bp product was purified by agarose gel electrophoresis followed by isolation using the Promega Wizard® SV Gel and PCR Clean-Up System. The purified *def-fmt* amplicon was cloned in to the pJET1.2 vector using the CloneJet PCR Cloning Kit (Thermo Scientific) as per the manufacturer's instructions and confirmed by sequencing (Macrogen). The resulting pJET::(*def-fmt*) plasmid was used to transform DF000. Expression was confirmed by RT-PCR and northern blotting.

Genome sequencing

Strains REL606 and DF000 were streaked to single colonies on DM2000 agar containing Streptomycin. Single colonies were used to inoculate DM2000 liquid media containing Streptomycin (for REL606) or Streptomycin, Kanamycin and Actinonin (for DF000). Genomic DNA was isolated using the Wizard Genomic DNA Purification Kit (Promega) and quantified using the Nanodrop 1000 Spectrophotometer and Qubit 2.0 Fluorometer. Sequencing was carried out by Macrogen Korea using an Illumina MiSeq platform with 2x250bp paired end reads. The resulting sequencing reads were processed using AdapterRemoval (Lindgreen, 2012) to remove low quality reads and adapter sequences. Reads were mapped to the REL606 genome (NC_012967) using Bowtie2 (Langmead and Salzberg, 2012) (using default parameters, specifying haploid genomes where necessary), and the mapped genomes visualized in Geneious v6.1.7 (Biomatters).

RT-PCR

RNA was isolated from early stationary phase cultures by the hot phenol method (Schmitt et al., 1990). Purified RNA was diluted to 200ng/µL and treated with TURBO DNase (Ambion) following the manufacturers guidelines. RT-PCR was carried out using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen) following the manufacturers guidelines. Primers used to detect *def-fmt* transcripts were:

*def-fmt*_fwd - 5′ – CGAGACGATGTACGCAGAAGAAGG – 3′

*def-fmt*_rev - 5′ – GAACTTCTGGTTTCGCCGTGCC – 3′

Primers used to detect *gstA* transcripts were:

*gstA*_fwd - 5′ – CTTTGCCGTTAACCCTAAGGG-3′

*gstA*_rev - 5′′ – GCTGCAATGTGCTCTAACCC – 3′.

Northern blotting

RNA was isolated from stationary phase cultures under acidic conditions as described previously (Köhrer and RajBhandary, 2008). Aminoacylated tRNA in the total cellular RNA sample was deacylated by the addition of Tris-HCl, pH 9.5 to a final concentration of 0.1M followed by incubation at 37°C for 1 hour. Non-formylated aminoacyl-tRNA was deacylated by the addition of $CuSO_4$ and sodium acetate, pH 5.0 to a final concentration of 10mM and 0.1M, respectively, followed by incubation at 37°C for 30 minutes. Approximately 400ng of total RNA was separated by electrophoresis using 6.5% polyacrylamide (19:1 acrylamide:bisacrylamide), 8M urea, 0.1M sodium acetate pH 5.0 at 300V for 48 hours in a cold room. The initiator tRNA consistently migrated between the xylene cyanol and bromophenol blue loading dye bands. A 10cm long portion of the gel containing the tRNA was transferred to a Hybond-N+ membrane (GE Healthcare Life Sciences) in 4x TE buffer (40mM Tris-HCl, 4mM EDTA, pH 8.0) at 10V for 16 hours. The membrane was then baked at 80°C for 2 hours. The membrane was washed twice with 6x SSC (1xSSC is 150mM NaCl, 15mM trisodium citrate, pH 7.0) followed by 4X SSC and 2x SSC.

$tRNA_i$ was probed using a 5' Digoxigenin (NHS-ester) labeled synthetic DNA probe (Integrated DNA Technologies) corresponding to bases 20-45 of the *E. coli* initiator tRNA (5′ – DIG-NHS-CTTCGGGTTATGAGCCCGACGAGCTA – 3′) and detected using Anti-Digoxigenin-AP Fab fragments (Roche) and CPD-*Star* (Roche) as per the manufacturer's instructions. Chemiluminescence was detected using a G:Box gel documentation system (Syngene).

Results

The def-fmt gene pair can be deleted from E. coli strain REL606

To examine whether *def* and *fmt* can be knocked out from the chromosome of REL606, a recombination-based 'gene gorging' procedure was used to replace the *def-fmt* gene pair, or the *fmt* gene alone, with an *nptII* gene cassette. The *nptII* gene encodes neomycin phosphotransferase which provides resistance to kanamycin, allowing selection of transformants with the knockout genotype arising from the 'gene gorging' process. The recombination procedure was carried out as shown in Figure 1. The resulting sequences of the *def-fmt* genomic region are shown in Supplementary Figures 1 & 2. The isolated REL606 Δ(*def-fmt*)::*nptII* strain was designated DF000 and the isolated REL606 Δ*fmt*::*nptII* strain was designated F000.

The recombination system sourced from phage lambda promotes recombination with as little as 40bp of near sequence identity. Due to the indiscriminate nature of this system, it was important to confirm that the *def-fmt* gene pair had indeed been removed from the chromosome of DF000, rather than the genes moving elsewhere in the chromosome as a result of an internal recombination event. To confirm the absence of *def* and *fmt* genes from DF000 was confirmed by RT-PCR. Primers internal to the *def-fmt* bicistronic mRNA were used. Transcripts resulting from the *def-fmt* operon were detectable in REL606, but not in DF000 (Figure 2), confirming successful removal of the *def* and *fmt* genes from the REL606 chromosome. Transcripts resulting from a control gene, *gstA*, were detected in all samples, confirming that the absence of RT-PCR products in our DF000 samples were not due to the absence of RNA, nor failure of the RT-PCR.

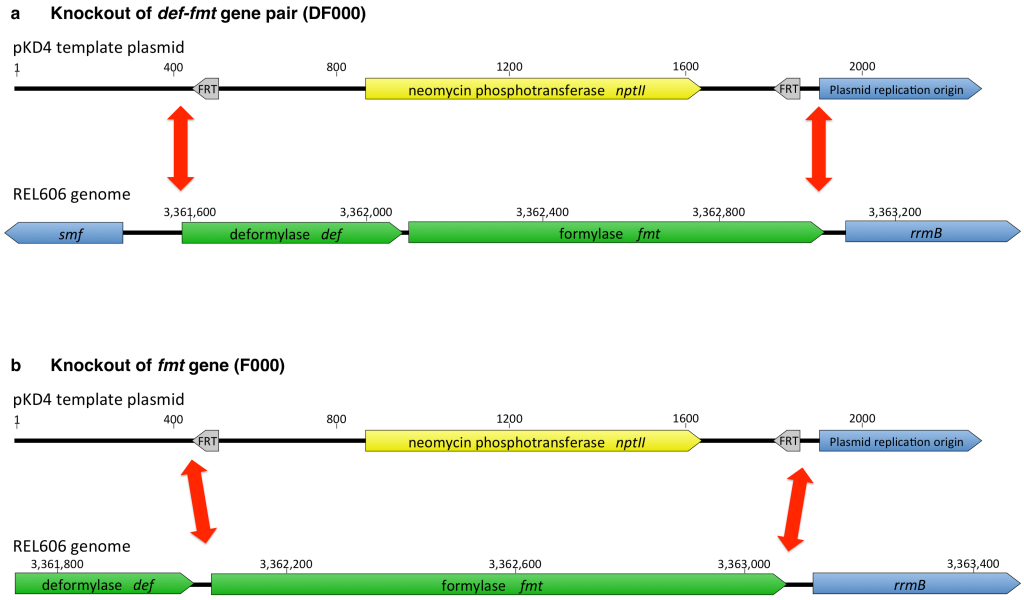


Figure 1 | Recombination of *def-fmt* region of REL606. a) DF000 was generated by replacing the *def-fmt* gene pair with an *nptII* cassette from pKD4. Sections of plasmid pKD4 and the REL606 genome are shown with numbers indicating base-pair position and labels indicating ORFs and sequence features. Red arrows indicate the recombination point. Recombination sites were designed immediately upstream of the *def* start codon and immediately downstream of the *fmt* stop codon. **b)** F000 was generated by replacing the *fmt* gene pair with an *nptII* cassette. Recombination occurred immediately upstream of the *fmt* start codong and immediately downstream of the *fmt* stop codon.

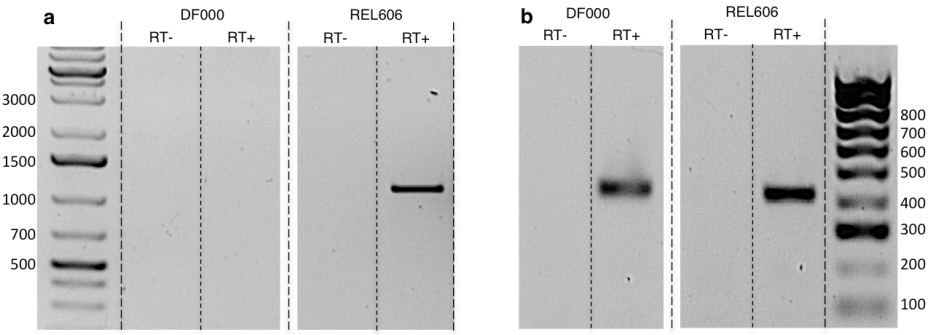


Figure 2 | DF000 does not produce *def-fmt* transcripts. Total RNA was extracted from early stationary phase cultures of DF000 and REL606 and used as a template for RT-PCR. RT- lanes were processed without a reverse transcriptase reaction in order to detect the presence of genomic DNA. RT+ lanes were processed as per the manufacturer's guidelines. **a)** RT-PCR with primers corresponding to REL606 *def-fmt* operon (expected size 1006bp). Lane 1, GeneRuler 1kb Plus (Thermo Scientific) with size standards indicated; lanes 2-5, RT-PCR products for DF000 and REL606 with RT- controls indicated. Positive amplification was only observed for REL606. **b)** RT-PCR with primers corresponding to *gstA* (expected product size 411bp). Lanes 1-4, RT-PCR products for DF000 and REL606 with RT- controls indicated; Lane 5, Hyperladder 4 (Bioline). Positive amplification observed for both DF000 and REL606.

DF000 does not formylate Met-tRNA$_i^{fMet}$

Knocking out the *def* and *fmt* genes from the *E. coli* chromosome does not confirm that formylation of the Met-tRNA$_i^{fMet}$ is no longer taking place. It is possible that *fmt* is dispensable because of redundant capacity provided by secondary activities of other enzymes. In order to determine if this were the case, we examined the tRNA$_i^{fMet}$ using northern blotting. This technique allows separation of fMet-tRNA$_i$, Met-tRNA$_i$ and tRNA$_i$ based on charge and size differences. Additionally, incubation of total cellular RNA at 37°C with 10mM $CuSO_4$ results in deacylation of aminoacyl-tRNAs with free *N*-termini e.g. Met-tRNA$_i$, but not of *N*-blocked aminoacyl-tRNAs e.g. fMet-tRNA$_i$. Incubation of total cellular RNA at 37°C at pH 9.5 results in deacylation of all aminoacyl-tRNA species. Accordingly, a Met-tRNA$_i$ formylating organism, e.g. *E. coli* strain REL606, should show a change in electrophoretic mobility of tRNA$_i$ following incubation at pH 9.5, but not after incubation with Cu^{2+}. In contrast, a non-formylating organism e.g. DF000 should show a change in electrophoretic mobility of tRNA$_i$ after both treatments. This can be seen in Figure 3, where a tRNA$_i^{fMet}$ band shift was detected by northern blotting after incubation at pH 9.5 (condition OH^-), but not after incubation with $CuSO_4$ (condition Cu^{2+}); a tRNA$_i^{fMet}$ band shift from DF000 was detected after both treatments (Figure 3b). The band shift observed after incubation of DF000 RNA with $CuSO_4$ indicates an unblocked N-terminus of Met-tRNAi, and hence the absence of detectable formylation. Expression of the *def-fmt* operon from the plasmid pJET::(*def-fmt*) in DF000 restores resistance of tRNA$_i^{fMet}$ to $CuSO_4$-derived deacylation (Figure 3c), that is, reversion to a wild-type formylation phenotype.

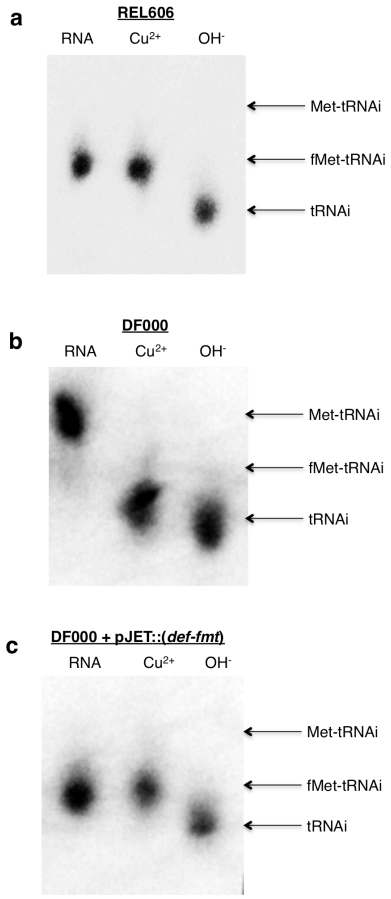


Figure 3 | DF000 does not formylate Met-tRNA$_i^{fMet}$. Total RNA from REL606, DF000 or DF000 containing plasmid pJET::(*def-fmt*) was isolated under acidic conditions. tRNA was either left untreated (RNA) or deacylated by incubation at pH 5 with 10mM $CuSO_4$ (Cu^{2+} condition) or pH 9 (OH^- condition) and examined by urea-PAGE followed by northern blotting for tRNA$_i^{fMet}$. **a)** tRNA$_i^{fMet}$ from REL606 shows no change in electrophoretic mobility following treatment with $CuSO_4$ indicating Met-tRNA$_i^{fMet}$ is formylated. Incubation of RNA at pH 9 results in a change in mobility of the tRNA$_i^{fMet}$ consistent with deacylation. **b)** tRNA$_i^{fMet}$ from DF000 shows a change in electrophoretic mobility following treatment with $CuSO_4$. This change in mobility is identical to incubation of RNA at pH 9, and is consistent with deacylation, indicating Met-tRNA$_i^{fMet}$ is not formylated. **c)** tRNA$_i^{fMet}$ isolated from DF000 expressing *def-fmt* from pJET::(*def-fmt*) shows resistance to deacylation by $CuSO_4$ similar to REL606. This suggests Met-tRNA$_i^{fMet}$ is formylated in this strain.

DF000 is resistant to actinonin

Formylation of Met-$tRNA_i^{fMet}$ in the absence of deformylation is lethal, and is thought to be the mechanism by which deformylase inhibitors function as antibiotics (Yuan et al., 2001). In the absence of formylation however, the peptide deformylase enzyme should have no substrate, hence bacterial strains which do not formylate Met-$tRNA_i^{fMet}$ should be resistant to peptide deformylase inhibitors such as actinonin. To examine the sensitivity of our strains to a peptide deformylase inhibitor, we grew REL606 and DF000 in varying concentrations of actinonin and measured their growth over time. REL606 was inhibited by actinonin at even the lowest concentration tested (5µg/mL), with complete inhibition of growth observed at 20µg/mL (Figure 4a). In contrast, DF000 could be cultured at all concentrations of actinonin tested, including 200µg/mL (Figure 4b). Growth was slowed by the highest concentration of actinonin (200µg/mL), however this growth inhibition may be attributable to the solvent used in production of actinonin stock solutions (DMSO). At an actinonin concentration of 200µg/mL, the final concentration of DMSO (2%) is sufficient to retard the growth of *E. coli* (Markarian et al., 2002).

Similar results were observed with REL606, DF000 and F000 plated on DM agar. REL606 showed very slight growth at 20µg/mL actinonin following 60 hours incubation at 37°C. In contrast, both DF000 and F000 were resistant up to and including 200µg/mL (Figure 5).

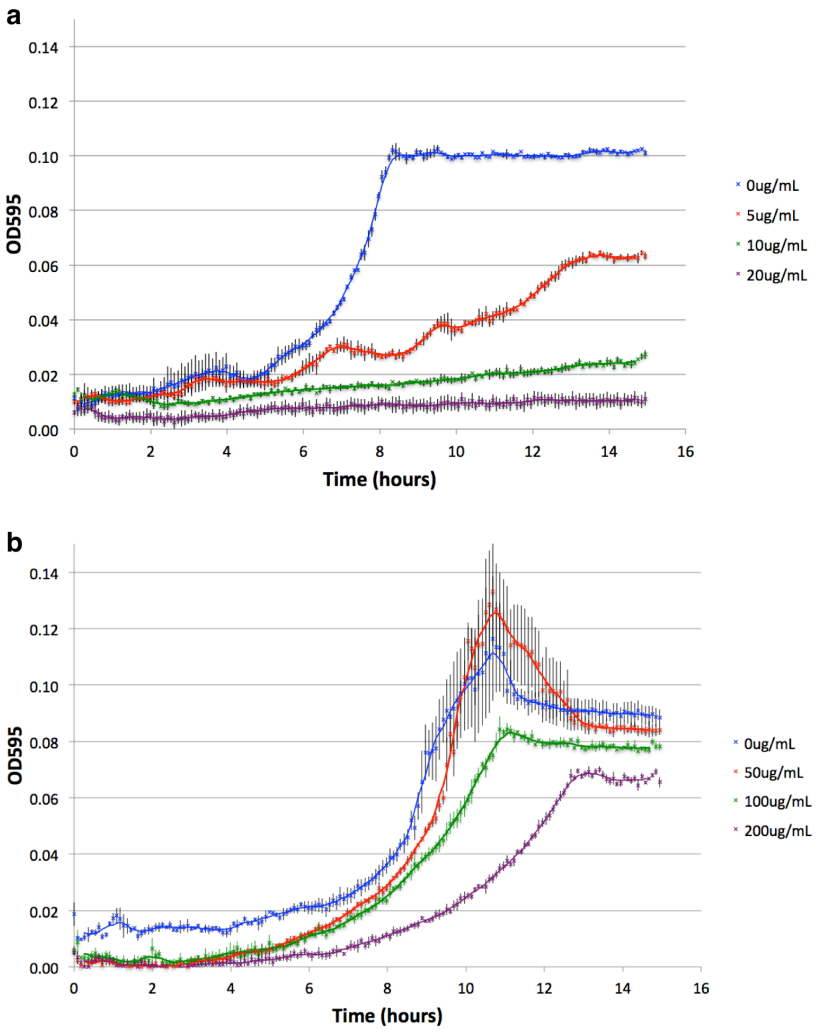


Figure 4 | DF000 is resistant to actinonin in liquid media. a) REL606 was grown in DM200 supplemented with increasing concentrations of actinonin (0, 5, 10, 20µg/mL indicated by blue, red, green and purple lines, respectively) and the growth of cultures measured by monitoring OD595. Growth was very slow at 10µg/mL and completely inhibited at 20µg/mL. **b)** Growth of DF000 was monitored in the presence of 0, 50, 100 and 200µg/mL actinonin (indicated by blue, red, green and purple lines, respectively). Note that these actinonin concentrations are 10-fold higher than those used for REL606. DF000 showed resistance up to 200ug/mL actinonin, albeit with slowed growth. Points represent $\overline{X}$±SD, n=3.

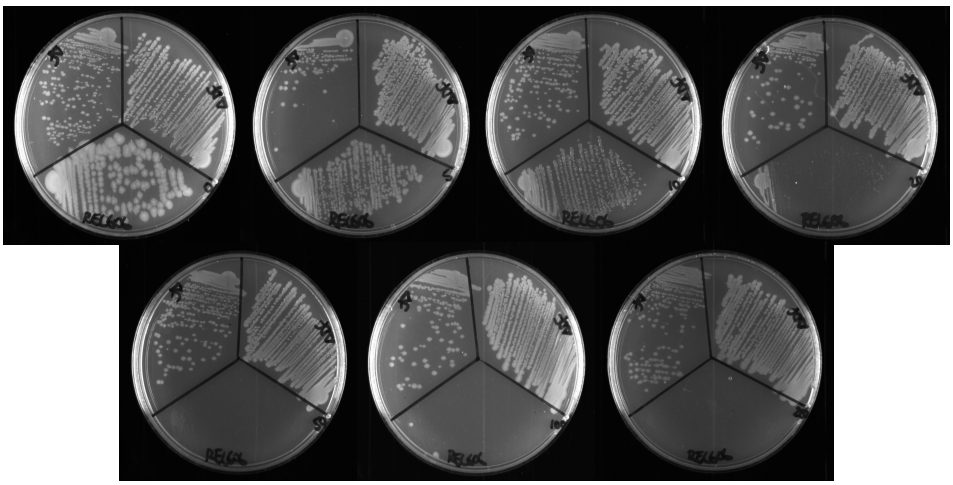


Figure 5 | *fmt* knockouts are resistant to actinonin on solid media. REL606, DF000 and F000 were grown on DM agar for 60 hours at 37C. Actinonin was supplemented at 0, 5, 10, 20 (top row, left to right), 50, 100 or 200µg/mL (bottom row, left to right). In all plates REL606 is bottom sector, F000 is top left sector and DF000 is top right sector. No growth of REL606 was visible above 20µg/mL actinonin, whereas DF000 and F000 show no inhibition of growth up to 200µg/mL actinonin.

DF000 grows more slowly than REL606

Removal of formylation activity from bacteria has been observed to reduce growth rate to varying extents (Chapter 1, Table 1). To determine whether knocking out the *def-fmt* gene pair slowed the growth of REL606, we measured the growth rate of DF000. Seven colonies of DF000 and three colonies of REL606 were isolated on DM2000 agar and grown to saturation in DM2000 liquid media before diluting 1:100 in fresh media to monitor growth. Surprisingly, these independent biological replicates grew at markedly different rates (Figure 6). Calculated minimum doubling times for the DF000 isolates increased to 95-140 minutes from 57 minutes calculated for the wild-type REL606. The appearance of faster growing mutants in our DF000 cultures may be due to compensatory mutations and is consistent with previous observations (Guillon et al., 1992; Mazel et al., 1994).

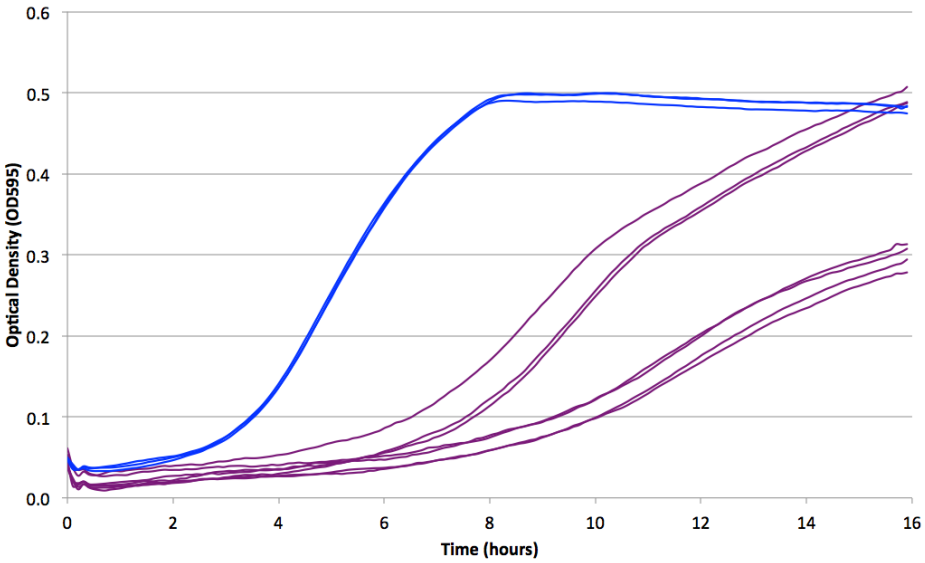


Figure 6 | DF000 grows more slowly than REL606. Three REL606 clones (blue) and seven DF000 clones (purple) were grown in DM2000 media at 37°C with shaking. The optical density (OD595) was monitored in a FLUOStar OPTIMA plate reader for 16 hours. REL606 clones show near identical growth, in contrast to DF000 clones which show variation in growth rates.

DF000 is viable at 42°C

It has been observed previously that knocking out the *fmt* gene from *E. coli* prevents growth at 42°C and above (Guillon et al., 1992; Mazel et al., 1994). To determine if our DF000 strain showed a similar growth impairment, we monitored the growth of REL606 and DF000 at 42°C. Individual colonies of REL606 and DF000 were isolated from DM2000 agar and used to inoculate DM2000 liquid medium. Cultures were grown at 37°C to stationary phase, before diluting 1:100 in fresh DM2000 medium. The optical density (OD595) of cultures was then monitored at 42°C for 25 hours (Figure 7). At this temperature, the growth of both REL606 and DF000 was slower than at 37°C with doubling times of approximately 67 and 245 minutes, respectively. In contrast to previous reports (Guillon et al., 1992; Mazel et al., 1994), our DF000 strain did grow at 42°C. As observed at 37°C, the growth rate of DF000 at 42°C differed between independent isolates. The difference in growth at elevated temperature between our DF000 strain and previously reported strains may be a media-dependent effect. Additionally, previous observations may not have measured growth for the 16-20 hours required to observe turbidity. Finally, the variation in growth rate between independent isolates of DF000 may be due to the appearance of mutations which compensate for the loss of formylation and facilitate growth at 42°C.

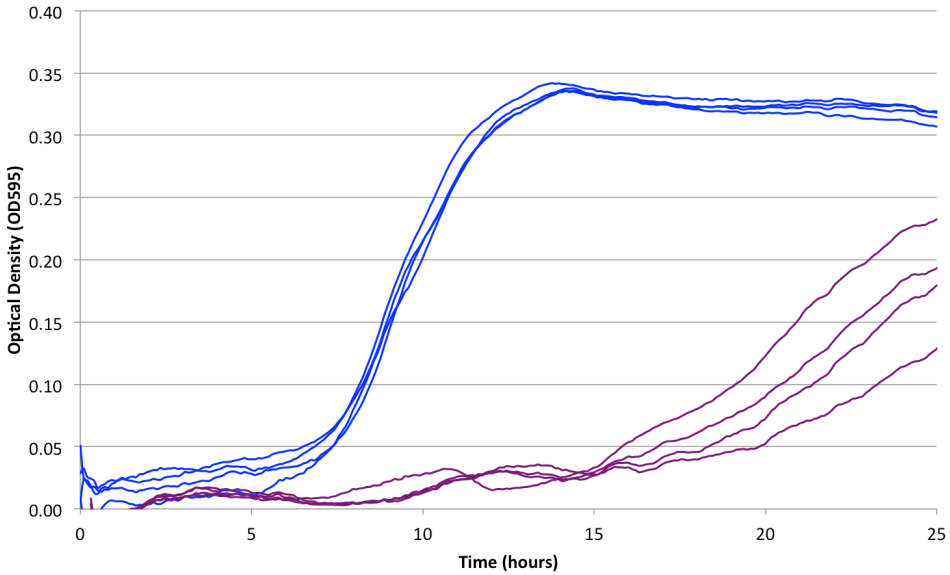


Figure 7 | DF000 is able to grow at 42°C. Four REL606 clones (blue) and four DF000 clones (purple) were grown in DM2000 media at 42°C with shaking. The optical density (OD595) was monitored in a FLUOStar OPTIMA plate reader for 16 hours. In contrast to previous reports, DF000 is able to grow at elevated temperature.

Discussion

The formylation of the methionyl-initiator tRNA is a process ubiquitous throughout the bacterial domain, but unique to bacterial translation (and translation in bacterial symbiont-derived organelles of eukaryotes (Giglione and Meinnel, 2001)). The genes encoding methionyl-tRNA formyltransferase (*fmt*) and peptide deformylase (*def*) have been identified in almost all sequenced bacterial genomes, indicating that this process has been conserved over billions of years. Despite this conservation, it has been known for over four decades that bacteria can initiate translation in the absence of formylation (Samuel and Rabinowitz, 1974; Samuel et al., 1970).

Here we show that formylation is not required in the *E. coli* strain REL606 to synthesise functional proteins. Knocking out the *def-fmt* operon from the genome of REL606 resulted in a strain with no detectable formylation of Met-$tRNA_i^{fMet}$. Our knockout strain, DF000, is resistant to the deformylase inhibitor, actinonin, at concentrations over 10-fold higher than wild-type cells. This resistance to actinonin suggests that the peptide deformylase enzyme does not function in important activities outside of formyl-peptide deformylation.

DF000 grew more slowly than its formylation-proficient parent strain REL606, consistent with previously isolated Δ*fmt* strains (Guillon et al., 1992; Mazel et al., 1994). In addition to growing more slowly, DF000 was much more variable in growth rate than REL606, with replicates derived from different colonies having divergent growth curves. Such variation has been described previously with formylation-deficient *E. coli* (Mazel et al., 1994) may be due to the appearance of mutations compensatory to the loss of formylation. The origin and function of these compensatory mutations will be investigated further in Chapter 3.

Supplementary Figure 1: Genomic sequence of DF000 *def-fmt* region.

Supplementary Figure 2: Genomic sequence of F000 *def-fmt* region.

References

Adams, J.M. (1968). On the release of the formyl group from nascent protein. J. Mol. Biol. *33*, 571–589.

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolić, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B., and Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature *452*, 108–111.

Chang, S.Y., McGary, E.C., and Chang, S. (1989). Methionine aminopeptidase gene of Escherichia coli is essential for cell growth. J. Bacteriol. *171*, 4071–4072.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. PNAS *97*, 6640–6645.

Giglione, C., and Meinnel, T. (2001). Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. Trends Plant Sci. *6*, 566–572.

Guenneugues, M., Caserta, E., Brandi, L., Spurio, R., Meunier, S., Pon, C.L., Boelens, R., and Gualerzi, C.O. (2000). Mapping the fMet-tRNAfMet binding site of initiation factor IF2. EMBO J. *19*, 5233–5240.

Guillon, J.M., Mechulam, Y., Schmitter, J.M., Blanquet, S., and Fayat, G. (1992). Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of Escherichia coli. J. Bacteriol. *174*, 4294–4301.

Herring, C.D., Glasner, J.D., and Blattner, F.R. (2003). Gene replacement without selection: regulated suppression of amber mutations in Escherichia coli. Gene *311*, 153–163.

Köhrer, C., and RajBhandary, U.L. (2008). The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases. Methods *44*, 129–138.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lindgreen, S. (2012). AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. BMC Res Notes *5*, 337.

Marcker, K., and Sanger, F. (1964). N-Formyl-methionyl-S-RNA. J. Mol. Biol. *8*, 835–IN838.

Markarian, S.A., Poladyan, A.A., Kirakosyan, G.R., Trchounian, A.A., and Bagramyan, K.A. (2002). Effect of diethylsulphoxide on growth, survival and ion exchange of Escherichia coli. Letters in Applied Microbiology *34*, 417–421.

Mazel, D., Pochet, S., and Marliere, P. (1994). Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO J. *13*, 914–923.

Meinnel, T., and Blanquet, S. (1995). Enzymatic properties of Escherichia coli peptide deformylase. J. Bacteriol. *177*, 1883–1887.

Petersen, H.U., Danchin, A., and Grunberg-Manago, M. (1976). Toward an understanding of the formylation of initiator tRNA methionine in prokaryotic protein synthesis. II. A two-state model for the 70S ribosome - Biochemistry (ACS Publications). Biochemistry.

Petersen, H.U., Joseph, E., Ullmann, A., and Danchin, A. (1978). Formylation of initiator tRNA methionine in procaryotic protein synthesis: in vivo polarity in lactose operon expression. J. Bacteriol. *135*, 453–459.

Samuel, C.E., and Rabinowitz, J.C. (1974). Initiation of protein synthesis by folate-sufficient and folate-deficient Streptococcus faecalis R. Biochemical and biophysical properties of methionine transfer ribonucleic acid. J. Biol. Chem. *249*, 1198–1206.

Samuel, C.E., D'Ari, L., and Rabinowitz, J.C. (1970). Evidence against the folate-mediated formylation of formyl-accepting methionyl transfer ribonucleic acid in Streptococcus faecalis R. J. Biol. Chem. *245*, 5115–5121.

Schmitt, M.E., Brown, T.A., and Trumpower, B.L. (1990). A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic Acids Res. *18*, 3091–3092.

Smith, A.E., and Marcker, K.A. (1970). Cytoplasmic methionine transfer RNAs from eukaryotes. Nature *226*, 607–610.

Solbiati, J., Chapman-Smith, A., Miller, J.L., Miller, C.G., and Cronan, J.E., Jr (1999). Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal. J. Mol. Biol. *290*, 607–614.

White, B.N., and Bayley, S.T. (1972). Methionine transfer RNAs from the extreme halophile, Halobacterium cutirubrum. Biochimica Et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis *272*, 583–587.

Wiser, M.J., Ribeck, N., and Lenski, R.E. (2013). Long-term dynamics of adaptation in asexual populations. Science *342*, 1364–1367.

Yuan, Z., Trias, J., and White, R.J. (2001). Deformylase as a novel antibacterial target. Drug Discovery Today *6*, 954–961.

Chapter 3

Formylation is not essential for wild-type growth rates in *E. coli*

All experimental work and subsequent analyses outlined in this chapter were performed by Ryan Catchpole under the mentorship of Prof. Jack Heinemann and Assoc. Prof. Anthony Poole. Bioinformatic analyses of genome sequences were performed by Ryan Catchpole with the assistance of Dr. Stinus Lindgreen. Protein modeling was performed by Katherine Donovan and Ryan Catchpole.

Introduction

The formylation of the methionyl-initiator tRNA (Met-$tRNA_i^{fmet}$) and the subsequent deformylation of peptides are activities that are present in all bacteria (Mazel et. al, 1997; Yuan et al., 2001). Despite its ubiquity, the formylation cycle can be removed from bacterial translation. In *E. coli*, this removal causes a significant decrease in the growth rate relative to formylating wild-type strains (Guillon et al., 1992; Mazel et al., 1994). A similar effect is also seen in the growth rate of other bacterial species, although the magnitude of the growth rate decrease differs between species (see Chapter 1, Table 1).

We have shown previously (see Chapter 2) that the *def-fmt* operon encoding the peptide deformylase and Met-$tRNA_i^{fMet}$ formyltransferase enzymes can be deleted from the genome of the *E. coli* strain REL606. The resulting strain, DF000, is able to carry out translation without the use of fMet-$tRNA_i^{fMet}$, albeit with a decreased

growth rate. We also noted that the growth rate of DF000 strains is often inconsistent between replicate cultures (Chapter 2, Figure 6) with faster growing mutants arising during culturing. This variation in growth rate has been observed previously (Guillon et al., 1992; Mazel et al., 1994) and may be due to the appearance of compensatory mutations arising in the slow growing *def-fmt* knockout strains.

We sought to investigate whether the growth rate of DF000 can be improved without the re-introduction of formylation and deformylation activity. We anticipated that the mechanism of growth rate recovery would be through the accumulation of mutations that compensate for the loss of formylation, however, the nature of such mutations is difficult to predict. Consequently, we used an experimental evolution approach whereby faster growing cells were able to arise through spontaneous mutation rather than through directed mutagenesis techniques. In our nutrient limited setup, these faster growing mutants would have a competitive advantage over their slower growing ancestors, growing to occupy a greater portion of the population prior to each re-culturing step. This allowed us to observe the accumulation of compensatory mutations through the selective advantage conferred by increased growth rate.

Methods

Long-term evolution

Single colonies of REL606 and DF000 were picked and grown to stationary phase in DM200. Each culture was diluted 1:100 in fresh DM200 and divided in to eleven 2mL replicate cultures and grown at 37°C for 24 hours. Each replicate culture was diluted 1:100 in fresh DM200 every 24 hours for 200 days. A control sample consisting of 2mL of DM200 media was incubated in parallel with the replicate cultures to test for sterility of media and aseptic transfer. After re-culturing, glycerol was added to each parental culture to a final concentration of 17% prior to storage at -80°C. Following every 7^{th} transfer, cultures were tested for contamination by visual examination of colonies on tetrazolium arabinose agar (Lenski et al., 1991; Levin et al., 1977), PCR screening of single colony isolates for the presence of the Δ(*def-fmt*)*::nptII* mutation, and sensitivity to T4 phage.

Determination of growth rate

Cultures were grown overnight in DM200 (or DM2000 where specified) and diluted 1:100 in fresh media. The OD_{595} of these cultures was then monitored for 16 hours at 37°C (with shaking) using the FLUOstar OPTIMA. The growth rate was determined as the minimum doubling time taken over a 30-minute interval.

Genome sequencing

Long-term cultures were streaked to single colonies on DM2000 agar containing streptomycin. Single colonies were used to inoculate DM2000 containing streptomycin (for REL606) or streptomycin, kanamycin and actinonin (for DF000). Genomic DNA was isolated using the Wizard Genomic DNA Purification Kit

(Promega) and quantified using the Nanodrop 1000 Spectrophotometer and Qubit 2.0 Fluorometer. Sequencing was carried out by Macrogen Korea using an Illumina MiSeq with 2x250bp paired end reads (2x300bp for temporal sequencing). The resulting sequencing reads were processed using AdapterRemoval (Lindgreen, 2012) to remove low quality reads and adapter sequences. Reads were mapped to the REL606 genome (NC_012967) using Bowtie2 (Langmead and Salzberg, 2012) (using default parameters, specifying haploid genomes where necessary), and the mapped genomes visualized in Geneious v6.1.6 (Biomatters).

RT-PCR

RNA was isolated from early stationary phase cultures by the hot phenol method (Schmitt et al., 1990). Purified RNA was diluted to 200ng/µL and treated with TURBO DNase (Ambion) following the manufacturers guidelines. RT-PCR was carried out using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen) following the manufacturers guidelines. Primers used to detect *def-fmt* transcripts were:

*def-fmt*_fwd - 5′ – CGAGACGATGTACGCAGAAGAAGG – 3′

*def-fmt*_rev - 5′ – GAACTTCTGGTTTCGCCGTGCC – 3′

Primers used to detect *gstA* transcripts were:

*gstA*_fwd - 5′ – CTTTGCCGTTAACCCTAAGGG-3′

*gstA*_rev - 5′ – GCTGCAATGTGCTCTAACCC – 3′.

Northern blotting

RNA was isolated from stationary phase cultures under acidic conditions as described previously (Köhrer and RajBhandary, 2008). Aminoacylated tRNA in the total cellular RNA sample was deacylated by the addition of Tris-HCl, pH 9.5 to a final concentration of 0.1M followed by incubation at 37°C for 1 hour. Non-

formylated aminoacyl-tRNA was deacylated by the addition of $CuSO_4$ and sodium acetate, pH 5.0 to a final concentration of 10mM and 0.1M, respectively, followed by incubation at 37°C for 30 minutes. Approximately 400ng of total RNA was separated by electrophoresis using 6.5% polyacrylamide (19:1 acrylamide:bisacrylamide), 8M urea, 0.1M sodium acetate pH 5.0 at 300V for 48 hours in a cold room. The initiator tRNA consistently migrated between the xylene cyanol and bromophenol blue loading dye bands. A 10cm long portion of the gel containing the tRNA was transferred to a Hybond-N+ membrane (GE Healthcare Life Sciences) in 4x TE buffer (40mM Tris-HCl, 4mM EDTA, pH 8.0) at 10V for 16 hours. The membrane was then baked at 80°C for 2 hours. The membrane was washed twice with 6x SSC (1xSSC is 150mM NaCl, 15mM trisodium citrate, pH 7.0) followed by 4X SSC and 2x SSC.

$tRNA_i$ was probed using a 5' Digoxigenin (NHS-ester) labeled synthetic DNA probe (Integrated DNA Technologies) corresponding to bases 20-45 of the *E. coli* initiator tRNA (5′ – DIG-NHS-CTTCGGGTTATGAGCCCGACGAGCTA – 3′) and detected using Anti-Digoxigenin-AP Fab fragments (Roche) and CPD-*Star* (Roche) as per the manufacturer's instructions. Chemiluminescence was detected using a G:Box gel documentation system (Syngene).

IF2 structural modeling

The IF2 crystal structure 1ZO1 was used as a template to build a homology model (SWISS-MODEL (Kiefer et al., 2009)) of the translated sequence of *infB* from REL606 containing the substitutions R749L, P807S and Y838C. Protein structures were visualized using PyMOL v1.3r1(DeLano, 2002).

MetZ promoter activity

The 150bp upstream of the *metZ* gene was synthesised 5′ to the EGFP gene and a ribosome binding site (Supplementary Figure 1). The construct was cloned in to the pUC57 vector. Site directed mutagenesis was then used to generate the SNPs observed in DF200 sequences (synthesis, cloning and mutagenesis carried out by Genscript USA Inc.). The resulting plasmids were used to transform REL606. Cultures were grown in DM2000 containing 100µg/mL ampicillin and fluorescence (ex. 485nm, em. 520nm) monitored during growth using a FLUOstar OPTIMA (BMG LabTech).

Results

The def-fmt gene pair is not required for wild-type growth rates

The *def-fmt* knockout strain DF000 was serially passaged in minimal media once daily for 200 days. The serial passaging was replicated with the ancestor of DF000, the wild-type strain REL606 in order to account for adaptation to the media or culturing conditions used. Each passage involved a 1:100 dilution of the culture in fresh medium, equating to 6.6 generations of growth per passage for a total of 1320 generations of serial culturing. An additional ~200 generations were necessary to isolate strain DF000, and to purify the end-point strains (named DF200), thus we estimate strain DF200 and REL606 diverged from a common ancestor approximately 1500 generations in the past.

During the course of our long-term evolution experiment (LTEE), the growth rate of wild-type REL606 control was seen to increase relative to the ancestral strain. This has been observed in previous LTEEs (Lenski et al., 1991) and is likely a result of adaptation to the media and culturing conditions.

The growth rate of our Δ(*def-fmt*) lines changed during the course of the LTEE (Supplementary Figure 2), increasing at a much greater rate than the wild-type control. The growth rate was indistinguishable from wild-type following approximately 1000 generations of culturing. As this growth rate increase is larger than the wild-type lines, it cannot be solely attributed to media/culturing adaptations and is likely a result of the accumulation of mutations compensatory to the loss of formylation activity. Serial passaging was continued to 1320 generations, during which time the growth rate did not change, remaining equivalent to wild-type (Figure 1).

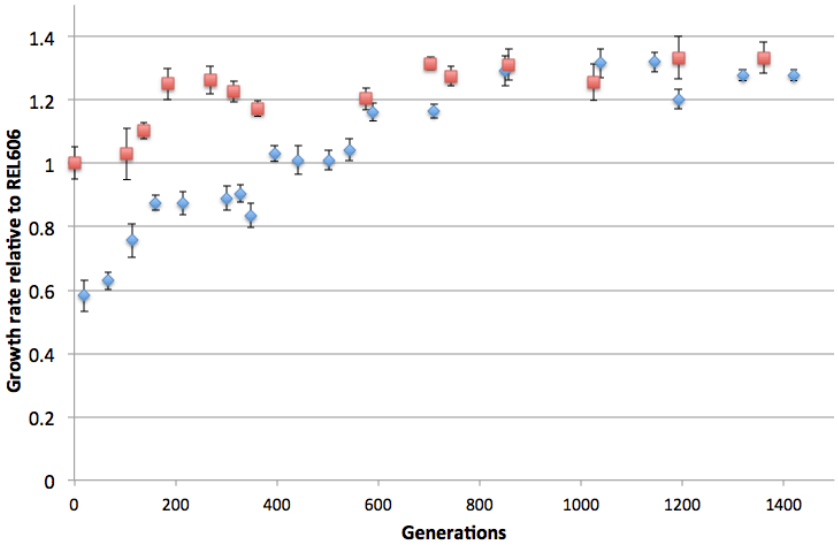


Figure 1 | Serial culturing of DF000 results in increased growth rate. *E. coli* strains DF000 and REL606 were serially cultured for 200 days. The growth rate was measured as the minimum doubling time over a 30-minute interval. The average growth rate (n=11, $\overline{X}\pm SE$) is shown relative to that of REL606, and plotted against the approximate generations of serial culturing. The growth rate of REL606 (red squares) increased to ~1.3x the ancestral rate following 1300 generations of culturing. The growth rate of DF000 (blue diamonds) was indistinguishable from wild-type following 1000 generations of culturing.

DF200 does not express def-fmt

Despite regular screening for contamination, it is possible that the growth rate compensation we observed was due to inadvertent introduction of formylase-proficient bacteria to our Δ(*def-fmt*) culturing lines, i.e. REL606 used as a LTEE control. To test for formylase and deformylase expression, we isolated total RNA from one growth rate compensated lineage (DF200) and used RT-PCR to test for the presence of *def-fmt* transcripts. We could not detect any transcription of *def-fmt* in DF200 (Figure 2). This is identical to the ancestral DF000 strain, and markedly different from REL606. Transcription of *gstA* was detectable in all samples indicating that the absence of *def-fmt* transcripts was not due to an absence of RNA. Isolates from the DF200 lineages were also resistant to actinonin at 200μg/mL, confirming the absence of wild-type formylation and deformylation in these strains. When taken in combination with previous contamination testing, this result shows that our DF200 lineage was not contaminated with wild-type *E. coli*. Additionally, this result shows that DF200 does not express the *def-fmt* gene pair.

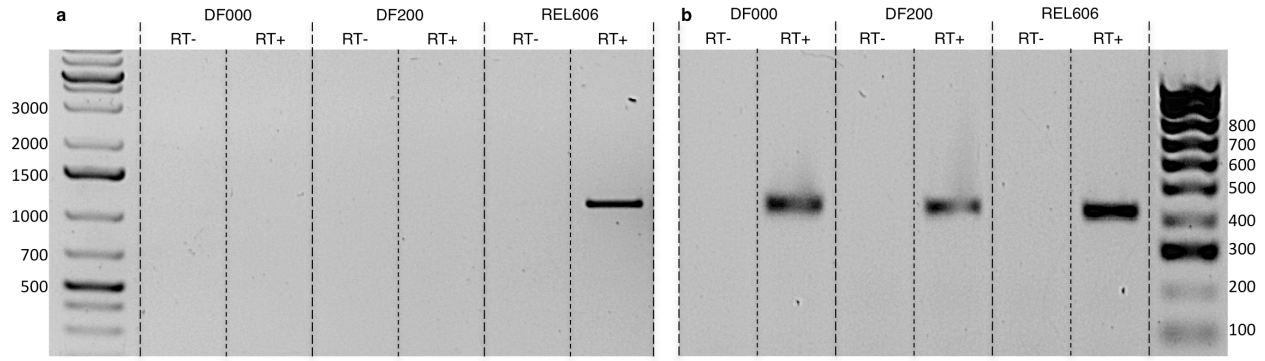


Figure 2 | No *def-fmt* transcripts detected in DF200. Total RNA was extracted from early stationary phase cultures of DF000, DF200 and REL606 and used as a template for RT-PCR. RT- lanes were processed without a reverse transcriptase reaction in order to detect the presence of genomic DNA. RT+ lanes were processed as per the manufacturer's guidelines. **a)** RT-PCR with primers corresponding to REL606 *def-fmt* operon (expected size 1006bp). Lane 1, GeneRuler 1kb Plus (Thermo Scientific) with size standards indicated; lanes 2-7, RT-PCR products for DF000, DF200 and REL606 with RT- controls indicated. Positive amplification was only observed for REL606. **b)** RT-PCR with primers corresponding to *gstA* (expected product size 411bp). Lanes 1-6, RT-PCR products for DF000, DF200 and REL606 with RT- controls indicated; Lane 5, Hyperladder 4 (Bioline). Positive amplification observed for all strains.

Met-tRNA$_i^{fMet}$ is not formylated in DF200

Although we could not detect transcripts of the *def-fmt* gene pair in DF200, it is possible that other cellular enzymes are carrying out formylation and deformylation in the absence of the *def-fmt* operon, leading to the increased growth rate of DF200. This process of promiscuous enzymatic activity is known as 'moonlighting' and has been observed previously as a mechanism to allow continued growth in the absence of a known enzyme (Copley, 2003; Näsvall et al., 2012). If an enzyme was moonlighting as a Met-tRNA$_i^{fMet}$ formyltransferase in our knockout strains, we would not expect to see a positive amplification product in RT-PCR, however we would expect to see a northern blotting signal consistent with fMet-tRNA$_i^{fMet}$. This was tested using northern blotting (Köhrer and RajBhandary, 2008). As described in chapter 2, fMet-tRNA$_i^{fMet}$ is sensitive to deacylation in basic conditions, but resistant to deacylation by $CuSO_4$ under acidic conditions. In contrast, non-formylated Met-tRNA$_i^{fMet}$ is sensitive to deacylation in both basic conditions, and in the presence of $CuSO_4$ in acidic conditions. Consequently, if tRNA$_i^{fMet}$ were isolated from a formylation-proficient organism, we expect to see no change in electrophoretic mobility following treatment with $CuSO_4$. In contrast, tRNA$_i^{fMet}$ is isolated from a formylation-deficient organism would be revealed through a change in electrophoretic mobility following treatment with $CuSO_4$. Examination of tRNA$_i^{fMet}$ isolated from DF200 shows that it was sensitive to deacylation by both $CuSO_4$ and incubation at pH 9 (Figure 3c), consistent with the presence of non-formylated Met-tRNA$_i^{fMet}$. In contrast, the blot detected a signal consistent with fMet-tRNA$_i^{fMet}$ from REL606 RNA - resistant to deacylation by $CuSO_4$, but sensitive to deacylation in basic conditions (Figure 3a). This confirms that strain DF200 did not have a formylated Met-tRNA$_i^{fMet}$, indicating translation must be initiated by non-formylated Met-tRNA$_i^{fMet}$.

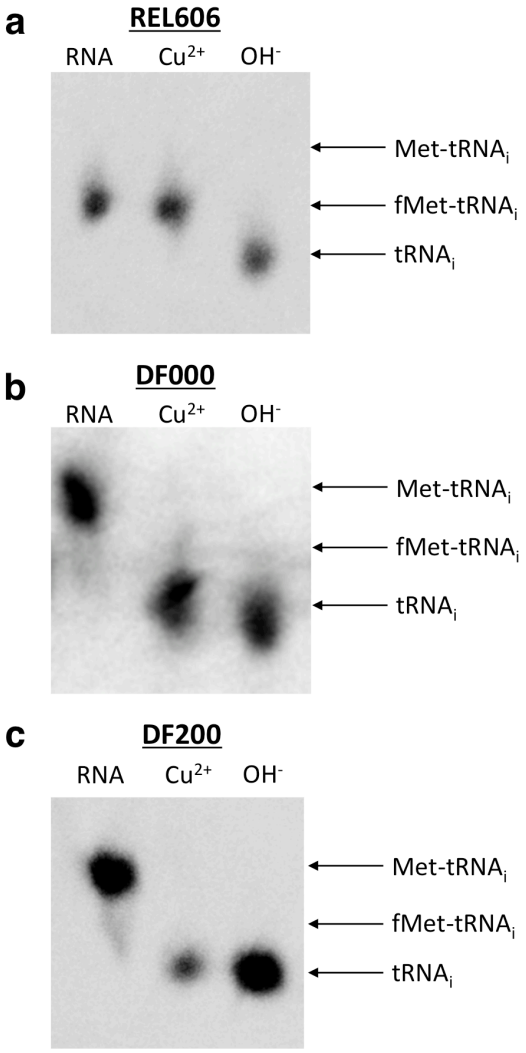


Figure 3 | DF200 does not formylate Met-tRNA$_i^{fMet}$. Total RNA from REL606, DF000 or DF200 was isolated under acidic conditions. tRNA was either left untreated (RNA) or deacylated by incubation at pH 5 with 10mM $CuSO_4$ (Cu^{2+} condition) or pH 9 (OH^- condition) and examined by urea-PAGE followed by northern blotting for tRNA$_i^{fMet}$. **a)** tRNA$_i^{fMet}$ from REL606 shows no change in electrophoretic mobility following treatment with $CuSO_4$ indicating Met-tRNA$_i^{fMet}$ is formylated. Incubation of RNA at pH 9 results in a change in mobility of the tRNA$_i^{fMet}$ consistent with deacylation. **b)** tRNA$_i^{fMet}$ from DF000 shows a change in electrophoretic mobility following treatment with $CuSO_4$. This change in mobility is identical to incubation of RNA at pH 9, and is consistent with deacylation, indicating Met-tRNA$_i^{fMet}$ is not formylated. **c)** tRNA$_i^{fMet}$ isolated from DF200 shows a deacylation pattern most similar to DF000. This suggests Met-tRNA$_i^{fMet}$ is formylated in this strain.

No deformylase activity in DF200

As peptide deformylase has no known substrate in the absence of formylation, so DF200 would derive no advantage from acquiring deformylation activity during the LTEE. However, an important but as yet unknown function of deformylase may promote the evolution of deformylation activity. The presence of deformylase activity in DF200 should allow growth in the presence of formylation, therefore, to test for the presence of deformylation in DF200, we attempted to introduce the *fmt* gene on a plasmid vector. Despite several attempts, we were unable to transform DF200 using an *fmt*-encoding plasmid. Moreover, expression of a plasmid-borne *def-fmt* operon in DF200 restored sensitivity to actinonin, indicating that formylation activity is toxic in the absence of deformylase activity in DF200. If DF200 had evolved deformylation activity in another cellular enzyme, the formylation-induced toxicity should be alleviated. These results suggest that DF200 is carrying out translation in the absence of both formylase and deformylase activity.

Genome sequencing reveals compensatory mutation in DF200

To examine the nature of the mutations responsible for growth-rate compensation in DF200, we sequenced the genomes of several isolates. Each of the eleven independent replicate lineages of DF200 and eleven wild-type control lineages, as well as the ancestral strains DF000 and REL606 were sequenced. All 24 samples were multiplexed in a single Illumina MiSeq lane, resulting in 6.4Gb of data across 27 million reads – an average of 265.3Mb±0.4Mb ($\overline{X}$±SD) of data per sample, giving an average sequencing depth of 59±9x per genome. The sequence data was of high quality, with 92% of bases having a Q30 value or higher.

Following cleaning of the reads with AdapterRemoval, they were mapped to either the REL606 genome (NC_012967.1), or a REL606-derived sequence containing the Δ(*def-fmt*)::*nptII* modification, using Bowtie2. The mapping gave a sequencing depth of 37±5x ($\overline{X}$±SD). Genotyping was then carried out using SNPest.

The DF200 strains acquired 50±15 ($\overline{X}$±SD) mutations during the LTEE compared to 4.6±1.7 in the control strain. The higher number of mutations in the knockout lineages is likely due to 10 out of the 11 DF200 lineages acquiring a mutator phenotype. A Luria–Delbrück Fluctuation Test (S E Luria, 1943) showed the mutation rate of strains with this mutation to be approximately 100 times higher than the wild-type lineages (data not shown). The source of this trait was identified as a $(GCTGGC)_3$ to $(GCTGGC)_2$ tandem repeat deletion in the mismatch repair gene *mutL*. This *mutL*-based mutator has appeared in previous LTEEs (Shaver and Sniegowski, 2003) and we might expect to observe an increased frequency of mutator alleles in our evolving lineages due to the strong selective pressure that a slow growth phenotype exerts (Denamur and Matic, 2006; Mao et al., 1997).

One of our independent lineages did not acquire a mutator phenotype by passage 200 (denoted DF200.1). The genome sequence of this line only showed 11 mutations relative to our ideal DF000 sequence. To determine whether these 11 mutations were sufficient to facilitate a growth rate identical to other lines, we measured the growth of DF200.1 alongside two other culturing lines – DF200.3 and DF200.6 (showing 53 and 58 mutations, respectively). Under these conditions, DF200.1 grew slower than DF200.3 and DF200.6 (Figure 4) indicating that this line has not accrued sufficient mutations to fully complement for the loss of formylation.

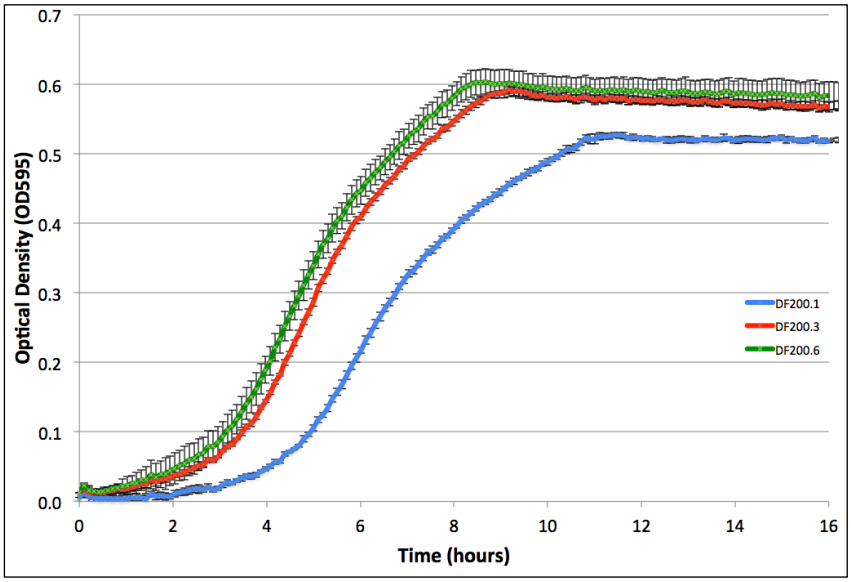


Figure 4 | DF200.1 grows more slowly than other DF200 lines. DF200.1 (shown in blue), DF200.3 (red) and DF200.6 (green) were grown to stationary phase in DM2000 media. Cultures were diluted 1:100 in fresh media and grown at 37°C with shaking. The optical density was monitored in a FLUOstar OPTIMA plate reader for 16 hours. DF200.1 grew slower than DF200.3 and DF200.6 during this time period.

Independent cultures have common mutations

The genome sequences of our independent DF200 lines show a remarkable degree of similarity in the mutations accrued during culturing (Figure 5). Comparison of the SNP profiles across the eleven lines revealed 35 mutations that arose in at least 6/11 lineages (Supplementary Table 1). The identical nature of these mutations suggests that our culturing lines have followed a similar evolutionary trajectory in overcoming the biochemical complications that a *def-fmt* knockout introduces. Several of the parallel mutations we observed in our DF200 lines were also present in our wild-type control lines. Polymorphisms in genes such as *pykF, spoT, nadR, rbs* operon etc. arose in both DF200 and our REL606-derived control lines and have also been observed previously in LTEEs using similar media conditions (Barrick et al., 2009). As a result, these SNPS can be attributed to adaptation to the media and/or culturing conditions and were not considered as relevant to compensation to the absence of formylation.

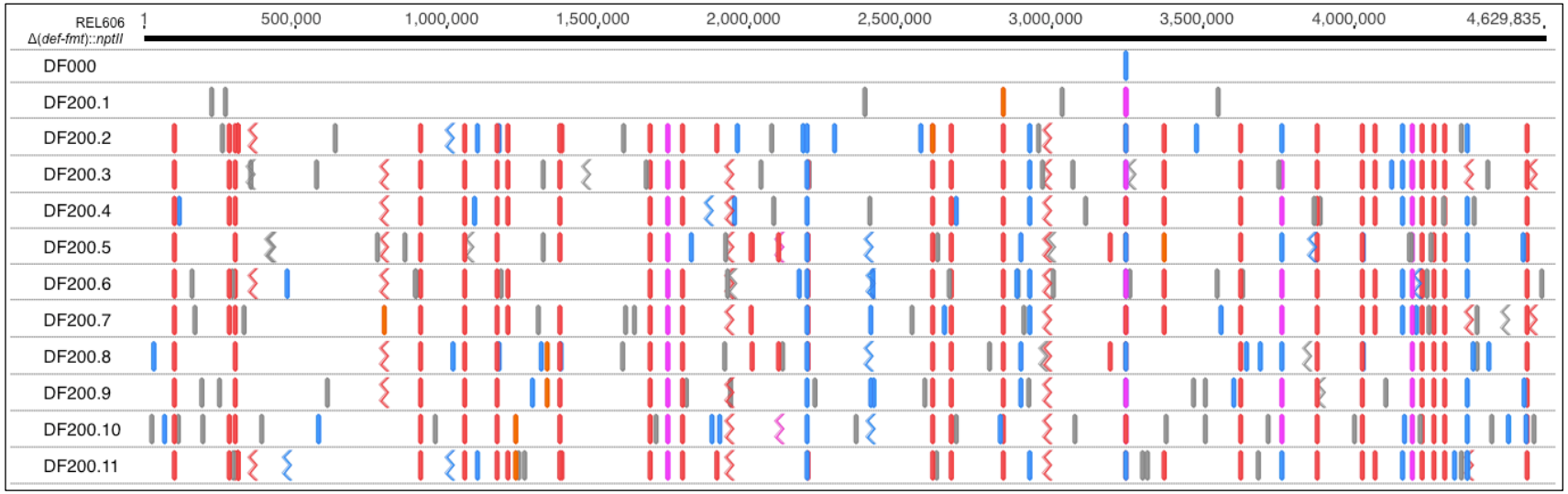


Figure 5 | Genome sequences of 11 independent long-term culturing lineages show parallel evolution. The genome sequence of the ancestral DF000 strain and 11 DF200 lines were mapped to the expected REL606 Δ(*def-fmt*)::*nptII* sequence using Bowtie2. Polymorphisms were detected using SNPest and are shown as vertical lines (SNPs) and zig-zag lines (indels). Polymorphisms are coloured to according to their similarity to other mutations: those in grey mutations unique to the DF200 line; red are identical to those observed in at least one other DF200 line; pink are in the same CDS as those observed in our control REL606 culturing lines; orange are in the same CDS as those observed in at least one other DF200 line; blue are in the same CDS as those observed in previous culturing experiments

We observed parallel mutation in several genes directly involved in translation. These genes may be important in compensating for the loss of formylation in translation initiation. Three of these are discussed further below.

SNPs upstream of MetZ drive increased $tRNA_i^{fMet}$ expression

Strikingly, all eleven DF200 lineages had a SNP in the region immediately upstream of the *metZ* operon, with 10 of 11 lines bearing an identical SNP. This operon contains three tandemly repeated copies of the initiator $tRNA_i^{fMet}$, thus mutations in this region of the chromosome may be important in regulating translation initiation. We noted that DF200.1 (the non-mutator lineage) of our LTEE had a C⇒T substitution 10bp upstream of *metZ* (denoted D1), whereas lineages DF200.2-11 (mutator lineages) had a C⇒A substitution 12bp upstream of *metZ* (denoted D2-11). These mutations are at positions -5 and -7 relative to the transcription start site (Nagase et al., 1988). It has been observed previously that increasing $tRNA_i$ levels partially alleviates the fitness cost of *fmt* mutations in *Salmonella enterica* (Nilsson et al., 2006). To assess whether the SNPs upstream of *metZ* impact $tRNA_i^{fMet}$ production, we developed a fluorescent reporter system using the 150bp region upstream of *metZ* to drive expression of GFP (Figure 6a). This reporter system was cloned into a plasmid vector and GFP fluorescence was monitored during the growth of REL606 strains containing our fluorescent reporter system. The GFP fluorescence resulting from D1 and D2-11 promoters was identical to wild-type during exponential growth, however the mutant promoters increase expression during early stationary phase relative to the wild-type promoter (Figure 6b,c).

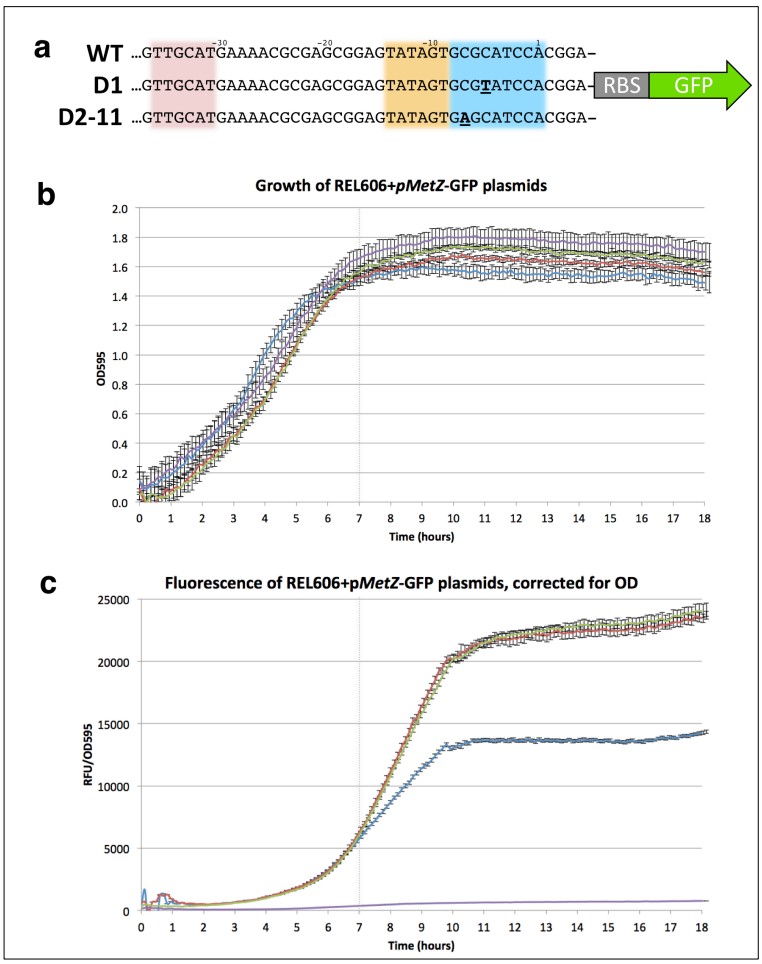


Figure 6 | SNPs in the *MetZ* promoter increase expression of downstream genes. a) A GFP reporter system, consisting of a plasmid-encoded GFP gene under the control of the wild-type (WT) or mutant *MetZ* promoters observed in our DF200 lines (D1, D2-11), was expressed in REL606. Nucleotides are labeled relative to the translation start site (Nagase et al., 1988) with the -35 sequence highlighted in pink and the -10 Pribnow Box in orange. Bold underlined nucleotides indicate the SNPs observed in DF200 lines. Blue highlighting indicates the region responsible for ppGpp downregulation of the promoter (Nagase et al., 1988) **b)** Growth of REL606 carrying the reporter plasmids (WT, blue; D1, red; D2-11, green), or an isogenic plasmid without the GFP gene (purple) was monitored at OD595 for 18 hours. **c)** The fluorescence of GFP (ex. 485nm, em. 520nm) in each culture was measured during growth. Fluorescence was corrected for culture turbidity by dividing relative fluorescence units (RFU) by OD595 measurements at each time point. The GFP fluorescence resulting from D1 and D2-11 promoters was indistinguishable from that of the WT promoter until cells begin entering stationary phase at 7 hours (vertical line), after which the fluorescence from D1 and D2-11 promoters is significantly higher than WT.

The results of our reporter assay reveal that both SNPs observed in the *metZ* promoter from our DF200 lines have the ability to increase gene expression, at least of a heterologous protein. It has been shown previously that the *metZ* promoter has a GC rich segment (spanning positions -8 to +1) responsible for downregulation of *metZ* transcription in the presence of ppGpp (Nagase et al., 1988). Both SNPs fall in this region of the promoter, and both decrease GC content. Our results are therefore consistent with mutations in the *metZ* promoter facilitating increased transcription of $tRNA_i^{fMet}$, despite increasing ppGpp levels as nutrients in the media become limiting. We predict that the increased production of $tRNA_i^{fMet}$ may help to overcome the low affinity that IF2 displays toward non-formylated initiator tRNA (Antoun et al., 2006).

Mutations in infB affect the fMet-tRNA$_i^{fMet}$ binding domain of IF2

Two SNPs were found to occur in *infB* during out LTEE, one of which occurred in 10 of the 11 parallel lines. This mutation results in the amino acid substitution Y838C in translation initiation factor 2 (IF2), with P807S resulting from the second SNP. IF2 facilitates assembly of the 70S ribosome by positioning fMet-$tRNA_i^{fMet}$ in the P-site of the 30S ribosomal subunit and promoting docking of the 50S subunit. The tRNA binding domain (C2 subdomain) of bacterial IF2 is known to have a greater affinity for fMet-$tRNA_i$ than Met-$tRNA_i$, and IF2 has a greater subunit docking activity in the presence of formylated than non-formylated Met-$tRNA_i$ (Antoun et al., 2006).

To examine what effects the Y838C and P807S mutations might have on IF2 activity, we generated a homology model of the REL606 IF2 sequence based on the *E. coli* IF2 structure (1ZO1). We then mapped the positions of Y838C and P807S to the structure (Figure 7a). The *E. coli* IF2 structure 1ZO1 is based on an X-ray structure of IF2 from *Methanothermobacter thermoauotrophicus* fitted to Cryo-EM data from the *E. coli* initation complex (Allen et al., 2005). Consequently, it is

difficult to speculate on the precise function of our mutations as the exact positions displayed in our model may not be representative of a true initiation complex.

Both the Y838C and P807S mutations mapped to the C2 domain of IF2 (Figure 7c), responsible for recognition of the formylated methionyl-residue of $tRNA_i^{fMet}$ (Spurio et al., 2000). These mutations may alter the affinity of IF2 for Met-tRNAi, allowing efficient translation in the absence of fMet-$tRNA_i^{fMet}$.

A further SNP arose in *infB* of our ancestral DF000 strain, preceding the LTEE. This mutation resulted in the amino acid substitution R749L in the C1 domain of IF2. This substitution is adjacent to sites that have been shown to increase the propensity of IF2 to adopt the 50S docking conformation (Figure 7b). Such mutations allow the formation of 70S ribosomes regardless of the amino-acyl tRNA bound to IF2. Therefore, this mutation may facilitate ribosome assembly when Met-$tRNA_i^{fMet}$ is bound to IF2 in the absence of fMet-$tRNA_i^{fMet}$. This R749L substitution also maps to a similar region of IF2 as the D764E mutation observed in previous LTEEs, indicating that mutations in this region may also be important in adaptation to the culturing environment.

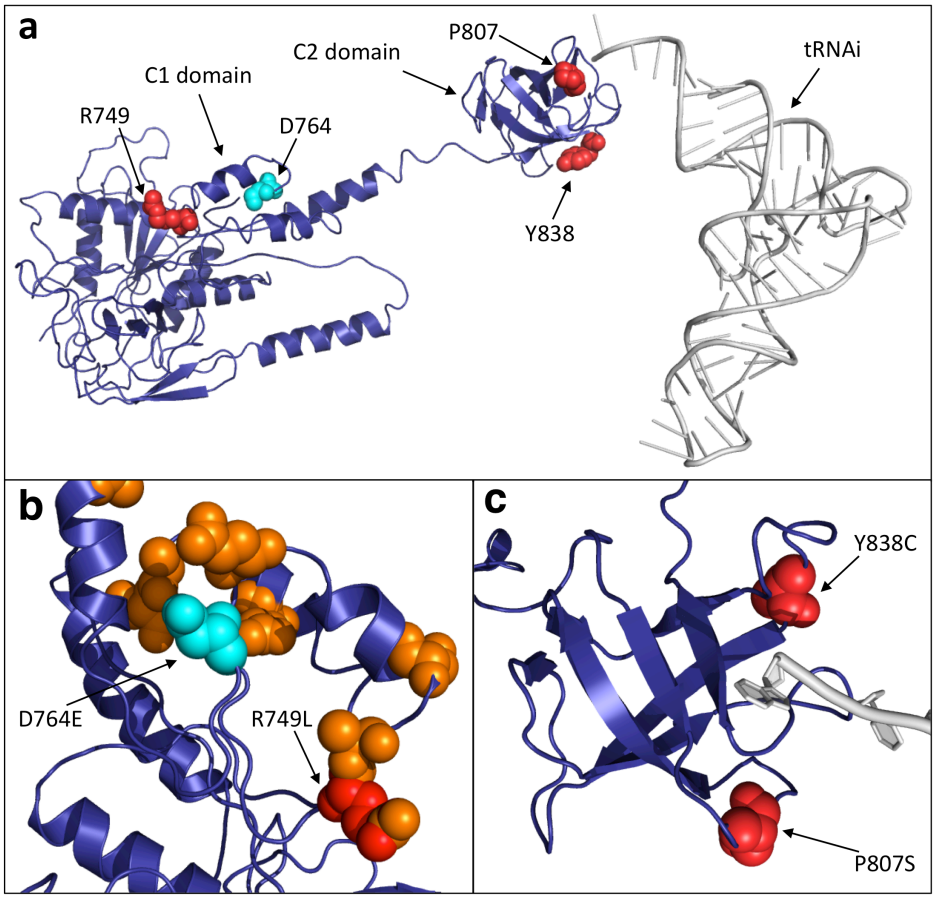


Figure 7 | Amino-acid substitutions in IF2 from DF200. a) The structure of IF2 from REL606 modeled on 1ZO1. IF2 is shown in purple and the $tRNA_i^{fMet}$ is shown in grey. The positions of mutations observed in DF200 lines are shown in red with the previously observed IF2 mutation shown in cyan. **b)** The C1 domain of IF2 showing the previously observed D764E substitution (cyan) and our R749L substitution (red). Residues shown in orange (including R749) are those where previously observed mutations have increased ribosome subunit docking. **c)** The C2 domain of IF2 with the DF200 mutations Y838C and P807S labeled in red. The 3′ terminus of $tRNA_i^{fMet}$ is shown in grey.

A mutation in rpmE2 may affect translation initiation

Seven of our 11 long-term culturing lines developed a SNP in the gene *rpmE2 (ykgM)*. This mutation results in a S50P substitution in the 50S ribosomal subunit L31. The *E. coli* genome encodes two genes for L31 – *rpmE* and *rpmE2*. As in other bacterial species, the *rpmE2* gene encodes an L31 protein which, unlike it's *rpmE* encoded counterpart, does not contain a CxxC zinc(II) binding motif (Nanamiya et al., 2004). In *Bacillus subtilis,* the *rpmE2*-equivalent gene *ytiA* is expressed under conditions of zinc starvation and due to it's higher affinity for the ribosome, displaces the canonical *rpmE*-encoded L31 protein from the 50S ribosomal subunit (Akanuma et al., 2006). This process is thought to release Zn(II) from L31 of assembled ribosomes and to limit further Zn(II) use in the cell by providing alternative L31 protein which does not require Zn(II). The medium used in our LTEE contained no zinc, hence *rpmE2* is likely to be the main L31 protein used in the ribosomes of DF200, and any functional mutations in L31 are likely to be seen in *rpmE2* rather than *rpmE.* A sequence alignment of *rpmE*-derived L31 and *rpmE2*-derived L31 from REL606 shows the S50P mutation aligns to the first cysteine of the CxxC motif.

The function of L31 is unclear. The protein is weakly associated with ribosomes and is sensitive to oxidation and degradation, thus it can often be lost during ribosome purification and structure determination (Eistetter et al., 1999). There is evidence to suggest that ribosomes containing *rpmE2*-derived L31 exhibit slower translation initiation than those containing the paralogous *rpmE*-derived L31 (Hindrimae 2013). Additionally, ribosome crosslinking experiments suggest that L31 interacts with tRNA in the ribosomal P-site (Bausk et al., 1984) and has a role in tRNA proofreading (Jenner et al., 2010). Thus, the mutation observed in *rpmE2* may play an important role in translation initiation in the absence of formylation.

A crystal structure has been solved for several bacterial ribosomes, however, the position of L31 in the ribosome is unclear. L31 is often built into ribosome structures adjacent to L9 (as in the *E. coli* structure 2AW4) (Schuwirth et al., 2005). This position places L31 deep in the ribosome, restricting its structure to one where the polypeptide folds back upon itself (Figure 8). When the structure of the *Thermus thermophilus* 70S ribosome was solved (structures 2J00 and 2J01), it was noted that the canonical position of L31 (as in 2AW4) contained electron density that was not attributable to L31. Instead, this density corresponded better with L28, and a second area of unknown electron density was attributed to L31 (Selmer et al., 2006). This new L31 position fits better with biochemical studies, and shows an unstructured L31 protein making contacts with both the 50S and 30S subunits (Figure 8). In this unstructured state, the S50P mutation is unlikely to affect the overall structure of the protein, however, due to the restrictive ring conformation of proline, the flexibility of the peptide chain may be reduced.

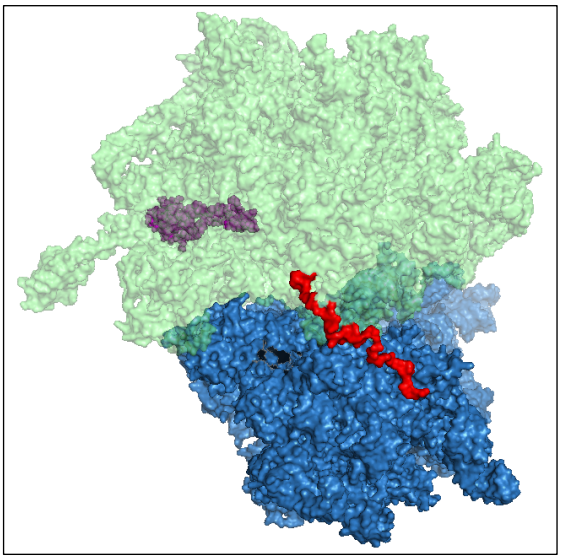


Figure 8 | Revised position of L31 in 70S ribosome. The original L31 position is occupied by L28 (purple) in the *T. thermophilus* ribosome structure, and the position of L31 (red) has been revised to the outside of the ribosome, making contact with the 50S (transparent green) and 30S (blue) subunits.

Multiple mutations are required to restore the growth rate of DF000

The evolution of mutations in the same gene, or in the same base position in multiple independent culturing lines suggests these SNPs may confer an adaptive advantage to our DF200 lines. To examine which, if any of these mutations confer a growth advantage to our knockout strain we measured the growth rate of one of our evolved lines (DF200.6) every 5 transfers from DF000-DF200. We then isolated genomic DNA from each of these time points and used next-generation sequencing to look for the appearance of specific alleles that correlate with increases in growth rate.

It is clear from the genome sequence data and the growth rate data (Figure 9) that no single mutation is sufficient to restore the growth rate of def-fmt knockout strains to wild-type. The doubling time of our DF000-derived culture decreases as the number of mutations increases through time, however it is difficult to attribute any single mutation to a significant change in growth rate. The simultaneous appearance of multiple mutations, combined with the error in growth rate measurements hampers our ability to find such influential mutations. Additionally, the sequencing reads obtained from our cultures (data not shown) suggest that most time points are composed of mixed populations of cells comprising both wild-type and mutant sequences at individual SNP locations. This heterogeneous population may limit our ability to resolve time points where the growth rate is representative of SNP-containing individuals. Despite this, our data confirm that compensation for the loss of formylation is not the result of a simple mutational change, but rather the combined effect of multiple individual mutations.

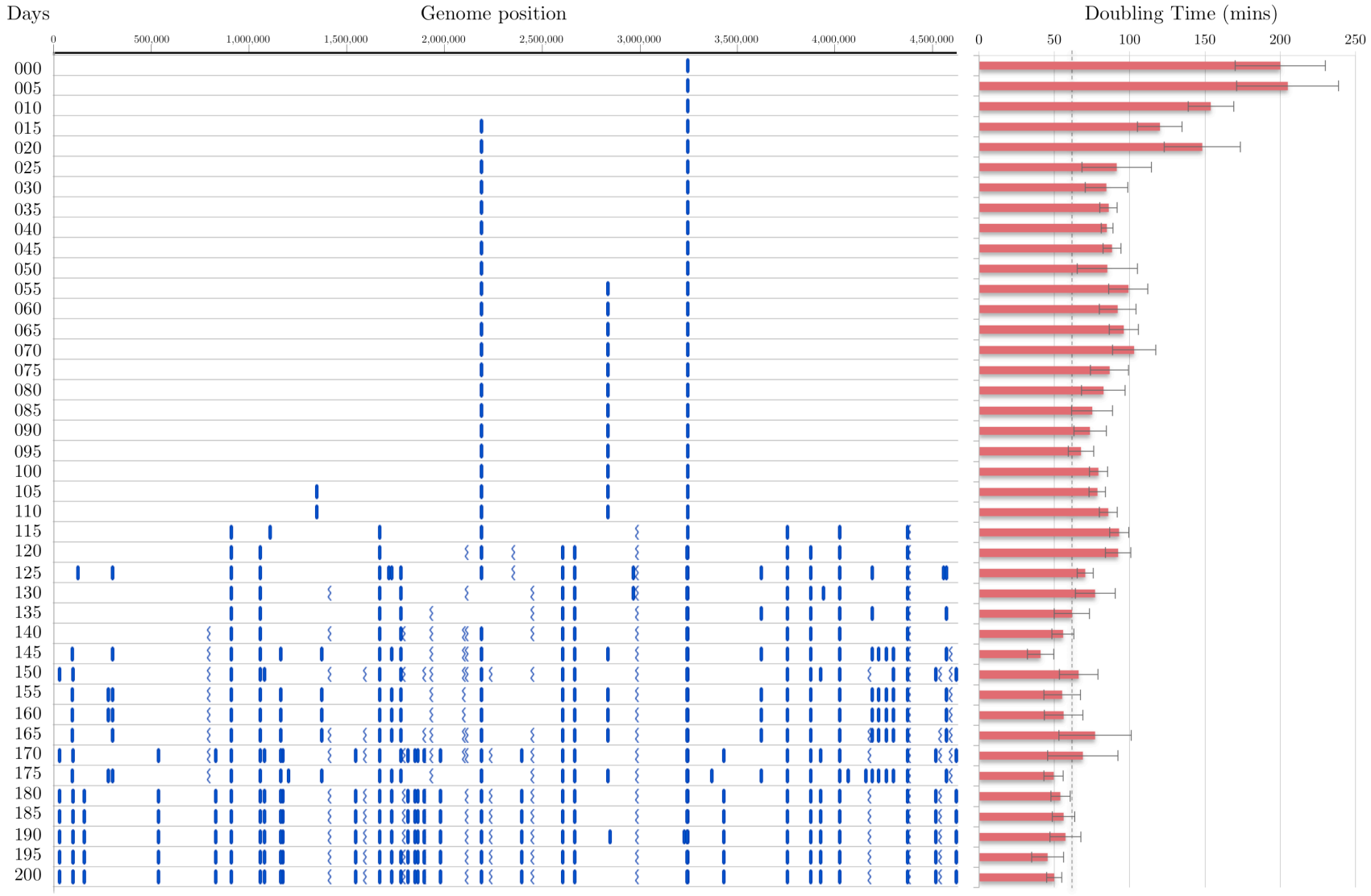


Figure 9 | Multiple mutations are required to compensate for loss of *def-fmt*. DNA was isolated at 5-day intervals during the long-term culturing of strain DF000. The DNA was sequenced by next-generation sequencing and mapped to the ideal DF000 genome (REL606 Δ(*def-fmt*)::*nptII*. Days of serial-culturing are shown on left, SNPs are indicated as vertical blue lines relative to position in DF000 genome, indels are shown as zig-zag lines. The doubling time of cultures was also measured at 5-day intervals during long-term culturing (red bars, $\overline{X}$±SE, n=14). Vertical dashed line indicates REL606 growth rate.

The appearance of the *mutL* mutator allele (discussed above) at day 115 correlates with the appearance of multiple mutations throughout the genome. Interestingly, following the appearance of the mutator genotype at passage 115, several previously mutated genes acquired additional mutations. This occured in both the *metZ* promoter and in *infB* (discussed above). In the case of the *metZ* promoter, the SNP observed prior to appearance of the mutator allele is consistent with that observed in our DF200.1 line which did not develop the *mutL* tandem repeat deletion, i.e. the D1 SNP; the SNP observed following appearance of the mutator allele is consistent with the D2-11 SNP observed in all of our DF200 lines harbouring the *mutL* mutation (DF200.2-12). It is difficult to attribute any increase in growth rate to this mutational change as both mutant promoters appear to have identical activity (Figure 6) and the growth rate changes after passage 115 coincide with the appearance of multiple mutations.

A similar change in SNP position was observed in *infB*. Prior to passage 115 and appearance of the *mutL* mutator allele, the SNPs observed in *infB* result in the substitution mutations R749L and P625L. The sequence data from day 115 shows a mixture of the wild-type P625 allele, and the mutant L625 allele. By passage 120, the L625 allele is absent from the sequence data, having been replaced by wild-type P625. A further SNP appears by passage 120, with amino acid position 838 showing a mixture of wild-type UAC (tyrosine) and mutant cysteine (UGC) codons. This Y838C substitution becomes fixed by passage 125.

Table 1 | The appearance of a mutator phenotype correlates with changes in translation genes. Genome sequencing of populations at 5 passage intervals reveals the appearance of a tandem repeat deletion in *mutL* at day 115. The appearance of this mutator allele coincides with changes in the *metZ* promoter region and in two separate regions of *infB*. For all sequences shown, polymorphisms are shown in blue. Tandem repeats in *mutL* are indicated in red boxes. Promoter features of *metZ* are shown in orange (Pribnow box), blue (ppGpp responsive element) and green (*metZ* gene).

Passages since DF000	Sequence of *mutL*	Sequence upstream of *metZ*	Sequence of *infB*
REL606	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG...	TTCCGGT.....TACGTGC.....TCTACGA
000	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTACGA
005	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTACGA
010	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTACGA
015	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTACGA
020	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTACGA
025	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTACGA
030	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTACGA
035	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
040	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
045	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
050	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
055	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
060	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
065	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
070	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
075	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
080	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
085	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
090	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
095	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
100	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
105	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
110	CGAGCTGGCGCTGGCGCTGGCGC	AGTATAGTGCGTATCCACGGACGCGGGG	TTCTGGT.....TACTTGC.....TCTACGA
115	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGYATCCACGGACGCGGGG	TTCYGGT.....TACTTGC.....TCTACGA
120	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTRCGA
125	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
130	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
135	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
140	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
145	CGA------GCTGGCGCTGGCGC	AGTATAGTGAGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
150	CGA------GCTGGCGCTGGCGC	AGTATAGTGMGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
155	CGA------GCTGGCGCTGGCGC	AGTATAGTGAGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
160	CGA------GCTGGCGCTGGCGC	AGTATAGTGAGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
165	CGA------GCTGGCGCTGGCGC	AGTATAGTGMGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
170	CGA------GCTGGCGCTGGCGC	AGTATAGTGMGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
175	CGA------GCTGGCGCTGGCGC	AGTATAGTGAGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
180	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
185	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
190	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
195	CGA------GCTGGCGCTGGCGC	AGTATAGTGCGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA
200	CGA------GCTGGCGCTGGCGC	AGTATAGTGAGCATCCACGGACGCGGGG	TTCCGGT.....TACTTGC.....TCTGCGA

Discussion

The ability of *E. coli* to initiate translation without the formylation of Met-$tRNA_i^{fMet}$ was first shown by the inhibition of tetrahydrofolate production (a cofactor for methionyl-tRNA formyltransferase) by trimethoprim (Harvey, 1973). Since that time, genetic techniques have been used to show that, despite the ubiquity of *def* and *fmt* genes throughout the bacterial domain, formylation activity is not essential for translation in bacteria (Guillon et al., 1992; Mazel et al., 1997; Newton et al., 1999). We showed in chapter 2 that the *def-fmt* operon can be completely removed from the genome of *E. coli* strain REL606, leaving no detectable formylation of Met-$tRNA_i^{fMet}$. However, this loss of formylation comes at a cost; our *def-fmt* knockout strain, DF000, grows significantly slower than its otherwise isogenic parent strain REL606. Here, we have used experimental evolution to allow the accumulation of mutations that compensate for the slow growth phenotype caused by deletion of the *def-fmt* operon. We find that compensatory mutations sufficient to restore wild-type growth rate can be observed within 1500 generations of culturing.

The recovery of DF200 lines to growth rates indistinguishable from wild-type *E. coli* was not due to recovery of formylation activity (either through inadvertent introduction of *fmt*, or evolution of new formylase enzymes). Rather, the recovery of growth rate was due to the appearance of mutations in several genes involved in translation, as well as many genes with unknown functions.

Increased expression of the initiator $tRNA_i^{fMet}$ has been shown to partially alleviate the fitness cost of resistance to the deformylase inhibitor, actinonin. In *Salmonella enterica*, this increased expression was achieved through up to 40-fold duplication of the *metZW* gene pair (both *metZ* and *metW* encode $tRNA_i^{fMet}$) (Nilsson et al., 2006). All 11 of our DF200 lines had a SNP in the region immediately upstream of the *metZ* gene in a region known to have promoter activity. The mutant promoters

from our DF200 lines increased expression of a downstream GFP reporter gene relative to a wild-type promoter, however, higher levels of GFP were only observed in early stationary phase. The *metZ* promoter is downregulated by the stringent response signaling molecule (p)ppGpp through a GC rich nonanucleotide immediately upstream of the transcription start site (Nagase et al., 1988). Production of (p)ppGpp increases as nutrients become limiting (Dalebroux and Swanson, 2012), repressing *metZ* transcription during entry in to stationary phase, consistent with the downregulation of GFP expression observed with our wild-type promoter construct (Figure 6). The SNPs observed in our DF200 lines are in this (p)ppGpp responsive nonanucleotide and decrease the GC content. Consequently, GFP expression from the mutant promoter constructs continues to increase in early stationary phase. The increased in $tRNA_i^{fMet}$ expression may promote translation initiation in our formylase deficient *E. coli* by providing more Met-$tRNA_i^{fMet}$ substrate for IF2, overcoming the low affinity of IF2 for non-formylated Met-$tRNA_i^{fMet}$. Additionally, increased expression of *metZ* in early stationary phase may increase the $tRNA_i^{fMet}$ pool in cells, priming them for faster recovery from stationary phase following the next culturing passage.

Amino-acid substitutions in the C1 domain of initiation factor 2 have been observed to increase the propensity of IF2 to adopt a 50S docking conformation following binding to the 30S ribosomal subunit (Zorzet et al., 2010). This promotes assembly of 70S ribosomes, regardless of the tRNA bound to the C2 domain of IF2. In wild-type formylation-proficient cells, aberrant assembly produces a greater proportion 70S ribosomes unable to initiate translation, decreasing fitness. However, in formylation-deficient cells, this indeterminate ribosome assembly confers a fitness advantage, facilitating translation with non-formylated Met-$tRNA_i^{fMet}$ (Zorzet et al., 2010). Our DF000 strain developed a SNP in *infB* producing an R749L substitution in IF2. Both the position of this R749 residue and the arginine-to-leucine substitution are identical to substitutions shown to facilitate translation in the absence of fMet-$tRNA_i^{fMet}$ (Zorzet et al., 2010). We also observed

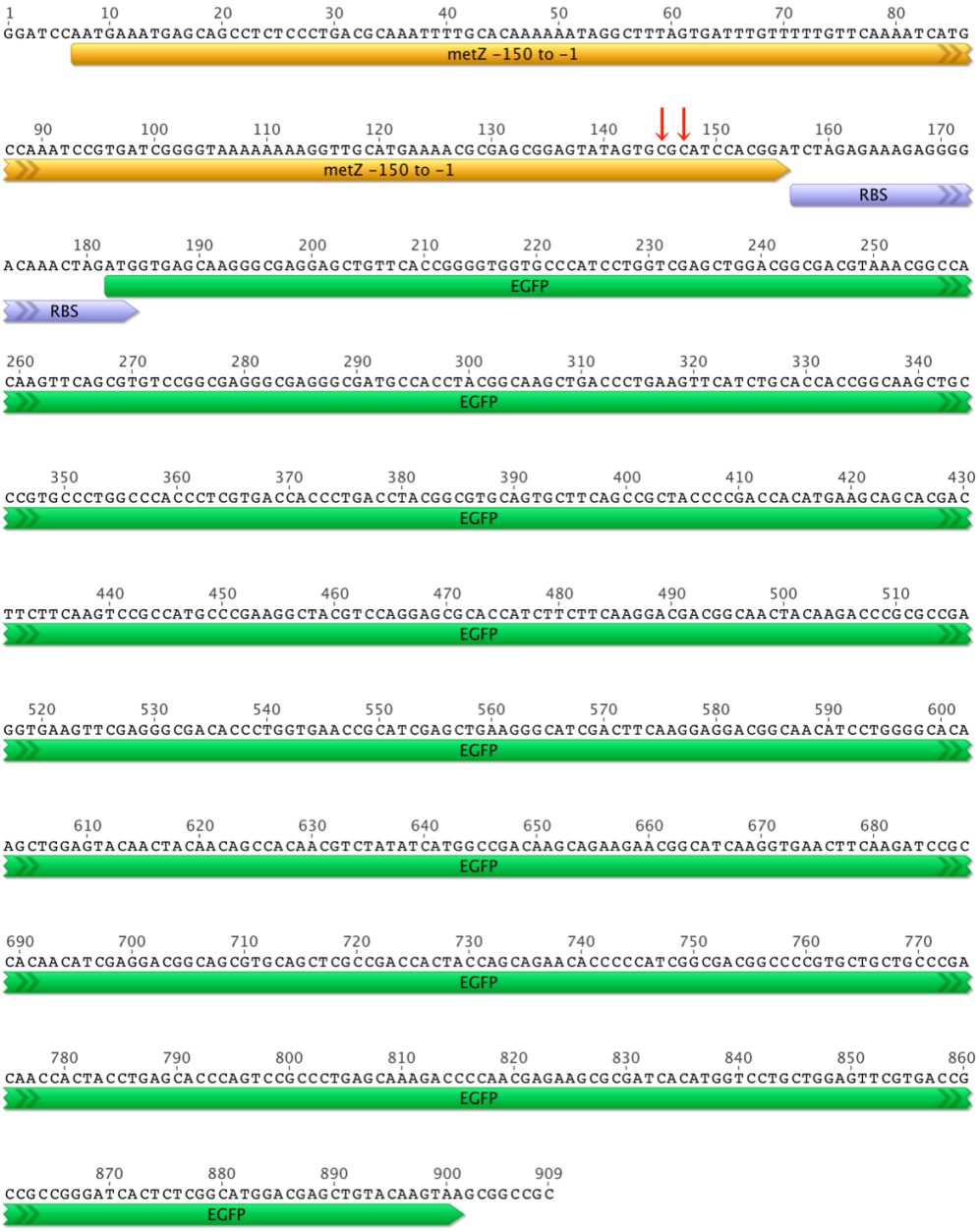
two further mutations in the fMet-tRNA$_i^{fMet}$ binding domain (C2) of IF2. These substitution mutations map to two seemingly flexible loops on either side of the fMet-tRNA$_i^{fMet}$ binding site. These substitutions may play a role in promoting non-formylated Met-tRNA$_i^{fMet}$ binding to IF2. It is difficult to speculate on the exact function that these substitutions play as our structural homology model is based on a low-resolution (13.8Å) cryo-electron microscopy structure of the *E. coli* IF2 in a 70S initiation complex (Allen et al., 2005), decreasing the structural accuracy of this inherently flexible molecule.

We also noted the appearance of mutations in several transmembrane proteins and membrane transport proteins e.g. *yeiB, betT, potH, yccC, malK, yjfD,* suggesting a functional role for formylation in membrane proteins. Indeed, it has been observed that a handful of *E. coli* membrane proteins (Coleman, 1992; Guan et al., 2011; Milligan and Koshland, 1990) as well as many inner membrane proteins in mitochondria retain the initiating *N*-formylmethionine residue. Moreover, bacteria secrete a small number of *N*-formylated peptide sequences into their environment. These data may indicate that formylation has evolved a role in membrane targeting of polypeptides. However, as with any other host-adaptive explanations for formylation, this function must be required only in bacterial cells (as well as bacterial endosymbiont-derived organelles), not in archaea or eukaryotes. A more likely explanation is found in the signal sequences targeting a protein to the cytoplasmic membrane; ribosomes synthesizing membrane proteins are targeted to the membrane before the *N*-terminal residue of the nascent polypeptide has emerged from the ribosome exit tunnel (Bornemann et al., 2008). Signal sequences encoded in the *N*-terminal residues of a protein are bound by signal recognition particles immediately upon their emergence from the ribosomal exit site. As a result, such peptides may not encounter the stoichiometrically limited deformylase before being bound by signal recognition particles and transported to the membrane. Consequently, membrane proteins may either retain the initiating fMet residue, or the fMet containing signal peptide may be cleaved during membrane

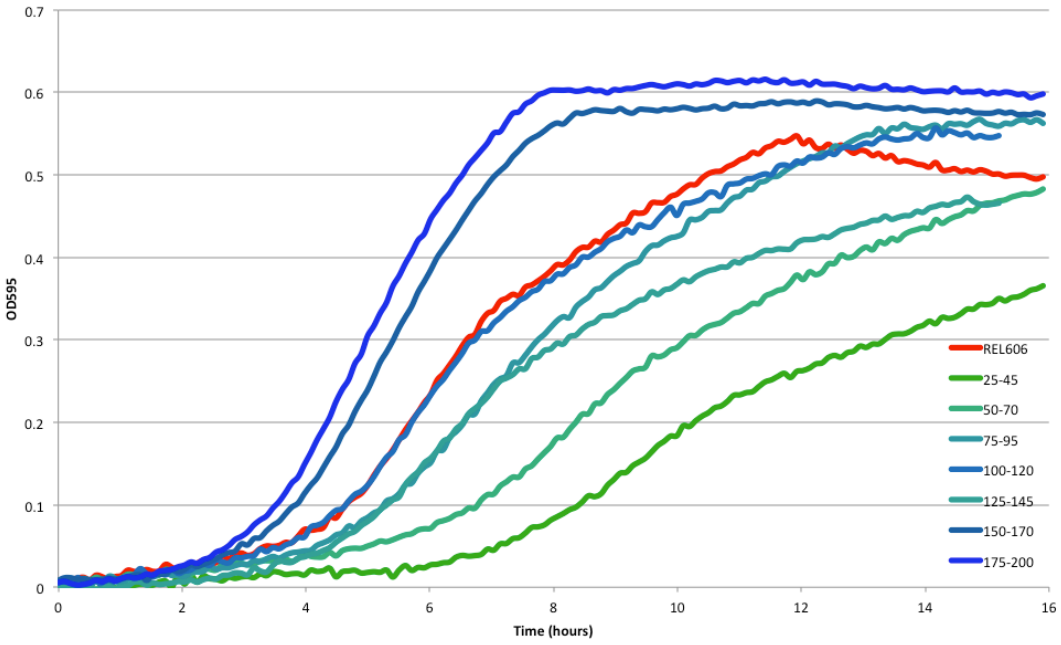
insertion resulting in the secretion of an fMet-peptide from the cell. The loss of formylation from our DF000 strain may alter the characteristics of proteins that usually preserve the fMet residue, requiring compensatory mutation to restore function.

Additionally, transmembrane proteins are often involved in the uptake of specific molecules from the environment. Due to the minimal nature of the media used during our culturing, we might expect to see a loss of function of some transport proteins.

It is clear from our data that formylation is not essential for wild-type growth rates in *E. coli*. However, multiple mutations were required to restore wild-type growth rates following the loss of formylation activity. This is important for understanding the persistence of this apparently dispensable activity in bacteria. In a natural competitive environment, any individual bacterium that loses formylation activity would experience a significant decrease in growth rate, and would be rapidly outcompeted by formylation proficient bacteria. Our data show that recovery of growth rate without formylation requires at least 1000 generations of isolation from wild-type bacteria (likely longer if wild-type mutation rates are maintained). Therefore such a recovery in growth rate may be near impossible in natural environments. We suggest that, rather than being essential to bacteria, the formylation and deformylation cycle may be too difficult to lose.



Supplementary Figure 1 | Synthesised sequence of *metZ* promoter and GFP gene. The sequence shown was synthesised by Genscript and cloned into the plasmid pUC57. Site directed mutagenesis was used to change the nucleotides at positions 145 and 147 (indicated by red arrows) to A and T, respectively. Orange arrow indicates identity to the 150bp upstream of *metZ* in REL606, purple indicates the synthetic ribosome binding site, green indicates the EGFP gene.



Supplementary Figure 2 | Growth curves measured during long-term culturing. Raw growth curves measured during long term culturing of DF000. Optical density measurements are shown as the average OD595 of 5 cultures isolated during the specified passage period.

Supplementary Table 1 | Mutations observed in 6 or more DF200 lineages

GENOME POSITION	GENE	DNA CHANGE	AA CHANGE	# LINES WITH MUTATION	GENE FUNCTION
99006	*mraY*	C -> T		10	Peptidoglycan synthesis
282810	*rpmE2*	A -> G	S50P	7	50S ribosomal protein L31
302927	*betT*	C -> T		10	Choline transport, osmoprotection
794719		$(G)_7$ -> $(G)_8$	94[†]	8	Hypothetical protein in prophage 434 (tail fiber)
913720	*potH*	A -> G	M294 V	10	Putrescine ABC transporter
1204233	*phoQ*	G -> A	P305L	6	Sensor protein (divalent cations, pH, acetate)
1374822	*ycjR*	A -> G	T141A	9	Sugar isomerase (unknown sugar)
1674102	*uidR*	T -> C	Y49C	10	Transcriptional regulator
1733896	*pykF*	T -> C	L311 P	10	Pyruvate kinase
1780023		$(A)_7$ -> $(A)_6$		9	Upstream of *thrS* (threonyl tRNA synthase)
1937937		$(C)_6$ -> $(C)_7$		8	Upstream of *cutC* (copper homeostasis, maybe P1 resistance)
2100307		deletion (2,100,308-2,122,430)		6	22kb deletion of P2-like prophage
2193373	*yeiB*	G -> A	A278V	10	Conserved inner membrane protein
2195257	*yeiG*	$(C)_9$ -> $(C)_8$	37[†]	7	Formaldehyde degradation
2609460	*yfhA*	G -> A	P15S	10	DNA binding response regulator. Respiration?
2670106		T -> C		10	Upstream of yjfD (inner membrane protein, Co^{2+} resistance?)
2841994		C -> A		10	*metZ* promoter
2930817	*bglA*	A -> G	Y468C	6	β-Glucosidase
2988292	*trmB*	$(C)_5$ -> $(C)_6$	74[†]	10	tRNA (m7G46) methyltransferase
3248736	*infB*	T -> C	Y838C	10	Initiation factor 2
3249003	*infB*	C -> A	R749L	11	Initiation factor 2 (PRESENT IN DF000)
3373150	rpsE	T -> C	H3R	6	30S ribosomal protein S5
3626578	*yhjS*	T -> C	V202A	10	Cellulose synthesis
3762755	*spoT*	A -> G	T659A	10	ppGpp synthase
3880988	*atpF*	T -> C		10	ATP synthase subunit
3895022	*rbsD-rbsA-rbsC-rbsK-rbsR*	deletion (3,895,023-3,900,128)		11	Ribose transport & catabolism (5kb deletion) PRESENT IN DF000
4028598	*engB*	A -> G		10	Cell division checkpoint
4071843		A -> G	N125S	6	Putative glycoporin
4161099	rpoB	G -> A	R74H	6	DNA-directed RNA polymerase subunit beta
4195400	*aceB*	A -> G	K152E	10	Malate synthase, glyoxylate shunt
4226334	*malK*	A -> G	T199A	10	Maltose ABC transport subunit
4266477		A -> G		10	Between *acs* (acetyl-CoA synthase) & *nrfA* (cytochrome c552)
4302715	*phnD*	A -> G	V159A	10	Phosphonate uptake
4375805	*mutL*	$(GCTGGC)_3$ -> $(GCTGGC)_2$	L(72)AR -> R	9	Mismatch repair
4575160	*hpaI*	G -> T	Q15K	10	Aldolase

References

Akanuma, G., Nanamiya, H., Natori, Y., Nomura, N., and Kawamura, F. (2006). Liberation of Zinc-Containing L31 (RpmE) from Ribosomes by Its Paralogous Gene Product, YtiA, in Bacillus subtilis. J. Bacteriol. *188*, 2715–2720.

Allen, G.S., Zavialov, A., Gursky, R., Ehrenberg, M., and Frank, J. (2005). The Cryo-EM Structure of a Translation Initiation Complex from Escherichia coli. Cell *121*, 703–712.

Antoun, A., Pavlov, M.Y., Lovmar, M., and Ehrenberg, M. (2006). How initiation factors maximize the accuracy of tRNA selection in initiation of bacterial protein synthesis. Mol. Cell *23*, 183–193.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243–1247.

Bausk, E.V., Graĭfer, D.M., and Karpova, G.G. (1984). Study of the photoaffinity modification of Escherichia coli ribosomes near the donor tRNA-binding center. Mol Biol (Mosk) *19*, 545–552.

Bornemann, T., Jöckel, J., Rodnina, M.V., and Wintermeyer, W. (2008). Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. Nature Structural & Molecular Biology *15*, 494–499.

Coleman, J. (1992). Characterization of the Escherichia coli gene for 1-acyl-sn-glycerol-3-phosphate acyltransferase (pIsC). Molec. Gen. Genet. *232*, 295–303.

Copley, S.D. (2003). Enzymes with extra talents: moonlighting functions and catalytic promiscuity. Curr. Opin. Chem. Biol. *7*, 265–272.

Dalebroux, Z.D., and Swanson, M.S. (2012). ppGpp: magic beyond RNA polymerase. Nature Reviews Microbiology *10*, 203–212.

DeLano, W.L. (2002). PyMOL molecular graphics system.

Denamur, E., and Matic, I. (2006). Evolution of mutation rates in bacteria. Mol. Microbiol. *60*, 820–827.

Eistetter, A.J., Butler, P.D., Traut, R.R., and Fanning, T.G. (1999). Characterization of Escherichia coli 50S ribosomal protein L31. FEMS Microbiology Letters *180*, 345–349.

Guan, Z., Wang, X., and Raetz, C.R.H. (2011). Identification of a chloroform-soluble membrane miniprotein in Escherichia coli and its homolog in Salmonella typhimurium. Anal. Biochem. *409*, 284–289.

Guillon, J.M., Mechulam, Y., Schmitter, J.M., Blanquet, S., and Fayat, G. (1992). Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of Escherichia coli. J. Bacteriol. *174*, 4294–4301.

Harvey, R.J. (1973). Growth and initiation of protein synthesis in Escherichia coli in the presence of trimethoprim. J. Bacteriol. *114*, 309–322.

Hindrimae, R (2003). Functions of Escherichia coli ribosomal protein L31 and L36 paralogues in protein synthesis. Unpublished Bachelors Thesis. University of Tartu, Estonia.

Jenner, L., Demeshkina, N., Yusupova, G., and Yusupov, M. (2010). Structural rearrangements of the ribosome at the tRNA proofreading step. Nature Structural & Molecular Biology *17*, 1072–1078.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. *37*, D387–D392.

Köhrer, C., and RajBhandary, U.L. (2008). The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases. Methods *44*, 129–138.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lenski, R.E., Rose, M.R., Simpson, S.C., and Tadler, S.C. (1991). Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. American Naturalist.

Levin, B.R., Stewart, F.M., and Chao, L. (1977). Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. American Naturalist.

Lindgreen, S. (2012). AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. BMC Res Notes *5*, 337.

Mao, E.F., Lane, L., Lee, J., and Miller, J.H. (1997). Proliferation of mutators in A cell population. J. Bacteriol. *179*, 417–422.

Mazel, D., Coïc, E., Blanchard, S., Saurin, W., and Marliere, P. (1997). A survey of polypeptide deformylase function throughout the eubacterial lineage. J. Mol. Biol. *266*, 939–949.

Mazel, D., Pochet, S., and Marliere, P. (1994). Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO J. *13*, 914–923.

Milligan, D.L., and Koshland, D.E. (1990). The amino terminus of the aspartate chemoreceptor is formylmethionine. J. Biol. Chem. *265*, 4455–4460.

Nagase, T., Ishii, S., and Imamoto, F. (1988). Differential transcriptional control of the two tRNA(fMet) genes of Escherichia coli K-12. Gene *67*, 49–57.

Nanamiya, H., Akanuma, G., Natori, Y., Murayama, R., Kosono, S., Kudo, T., Kobayashi, K., Ogasawara, N., Park, S.-M., Ochi, K., et al. (2004). Zinc is a key factor in controlling alternation of two types of L31 protein in the Bacillus subtilis ribosome. Mol. Microbiol. *52*, 273–283.

Näsvall, J., Sun, L., Roth, J.R., and Andersson, D.I. (2012). Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. Science *338*, 384–387.

Newton, D.T., Creuzenet, C., and Mangroo, D. (1999). Formylation is not essential for initiation of protein synthesis in all eubacteria. J. Biol. Chem. *274*, 22143–22146.

Nilsson, A.I., Zorzet, A., Kanth, A., Dahlström, S., Berg, O.G., and Andersson, D.I. (2006). Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. PNAS *103*, 6976–6981.

S E Luria, M.D. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics *28*, 491.

Schmitt, M.E., Brown, T.A., and Trumpower, B.L. (1990). A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic Acids Res. *18*, 3091–3092.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H.D. (2005). Structures of the bacterial ribosome at 3.5 A resolution. Science *310*, 827–834.

Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley,

A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. Science *313*, 1935–1942.

Shaver, A.C., and Sniegowski, P.D. (2003). Spontaneously arising mutL mutators in evolving Escherichia coli populations are the result of changes in repeat length. J. Bacteriol. *185*, 6076–6082.

Spurio, R., Brandi, L., Caserta, E., Pon, C.L., Gualerzi, C.O., Misselwitz, R., Krafft, C., Welfle, K., and Welfle, H. (2000). The C-terminal Subdomain (IF2 C-2) Contains the Entire fMet-tRNA Binding Site of Initiation Factor IF2. J. Biol. Chem. *275*, 2447–2454.

Yuan, Z., Trias, J., and White, R.J. (2001). Deformylase as a novel antibacterial target. Drug Discovery Today *6*, 954–961.

Zorzet, A., Pavlov, M.Y., Nilsson, A.I., Ehrenberg, M., and Andersson, D.I. (2010). Error-prone initiation factor 2 mutations reduce the fitness cost of antibiotic resistance. Mol. Microbiol. *75*, 1299–1313.

Chapter 4
The *def-fmt* operon exhibits a PSK phenotype

All experimental and bioinformatic analyses outlined in this chapter were performed by Ryan Catchpole under the mentorship of Prof. Jack Heinemann and Assoc. Prof. Anthony Poole. Bioinformatic analyses were performed by Dr. Stinus Lindgreen and Ryan Catchpole.

Introduction

In chapter 3 we showed that formylation of the methionyl-initiator tRNA (and subsequent peptide deformylation) is not required for wild-type growth rates in *E. coli*. Despite the apparent superfluity of formylation and deformylation, the genes encoding these activities have been identified in almost all known bacterial genomes (Yuan et al., 2001). These findings raise two important questions: 1) how could a dispensable function such as formylation spread to fixation in bacteria? And 2) how could formylation persist ubiquitously in bacteria over evolutionary time frames? We suggested in chapter 3 that a severe fitness disadvantage may prevent the accumulation of formylation-deficient individuals in bacterial populations. This may help to answer our question regarding persistence, however this is not sufficient to explain the fixation of formylation in bacteria.

The deformylase inhibitor, actinonin, is antibacterial due to the requirement for deformylation of peptides once formylation of Met-$tRNA_i^{fMet}$ has occurred (Mazel

et al., 1994; Yuan et al., 2001). Hence, the formylation of the initiator tRNA can be considered toxic. It then follows that the deformylation activity can be considered antidotal to this toxicity. This coupled toxin and antidote activity bears some resemblance to modern toxin-antitoxin gene pairs (TAs) that elicit a post-segregational killing phenotype (PSK) (Hayes, 2003; Van Melderen, 2010). TA gene pairs characteristically encode a stable toxin and labile antitoxin. Due to the difference in stability, the loss of a TA gene pair from a host cell results in accumulation of toxin relative to antitoxin, causing cell death (Van Melderen, 2010). TAs were first discovered by their ability to stabilize plasmids in bacterial populations by causing the death of plasmid-free segregants (Gerdes et al., 1986; Ogura and Hiraga, 1983). However, evidence suggests that this stabilization may be an artifact of monoculture conditions used in laboratory experiments (Cooper and Heinemann, 2000). Instead, the emergent property of toxin-antitoxin gene pairs may be competition between horizontally mobile elements (Cooper and Heinemann, 2000; 2005; Cooper et al., 2010; Kusano et al., 1995; Naito et al., 1995). When plasmids possess the dual characteristics of being horizontally mobile and able to elicit PSK, i.e. traits that allow competition between incompatible plasmids, a TA-encoding horizontally mobile plasmid can spread through a population of bacteria, displacing competing plasmids (Cooper and Heinemann, 2000; 2005; Cooper et al., 2010). In this way, TAs are able to successfully spread through bacterial populations, despite conferring a cost to hosts (Rankin et al., 2012). Nevertheless, the initial experiments used to identify TAs that cause PSK serve as a reliable diagnostic of TA systems that have evolved as PSK systems.

In addition to plasmids, TAs are found in almost all bacterial genomes (Pandey and Gerdes, 2005) where they have been implicated in many important functions. These include programmed-cell death (Engelberg-Kulka et al., 2004), stress response to nutrient limitation (Gerdes et al., 2005), genomic stability (Szekeres et al., 2007) and abortion of phage infection (Hazan and Engelberg-Kulka, 2004; Pecota and Wood, 1996). Furthermore, the presence of TAs in the bacterial

genome abolishes competition between horizontally mobile elements by facilitating continued antitoxin production following loss of a plasmid-borne TA (Cooper and Heinemann, 2005).

The *def-fmt* gene pair satisfies the two major prerequisites for PSK activity:

1) Toxin and antitoxin activity: for a TA to operate, one gene of the pair must encode a toxic activity and the other an antidotal activity. As discussed previously (see Chapter 1), bacteria cannot tolerate a formylated proteome, thus formylation can be considered toxic in the absence of deformylation. Deformylation relieves this toxicity, so can be considered an antitoxin to formylation. Thus the *def-fmt* gene pair appears to satisfy this toxin-antitoxin condition.
2) Lability of the antitoxin activity relative to the toxin: PSK requires the antitoxin activity to degrade faster than the toxin activity or toxic product per se, thereby causing an imbalance of toxin and consequent cell death or stasis. Previous efforts to purify and characterize the deformylase enzyme were hampered by low stability(Adams, 1968; Livingston and Leder, 1969; Rajagopalan et al., 1997) (activity of purified peptide deformylase has a half life of 60 seconds at room temperature), whereas the formylase enzyme is stable for months at room temperature (Blanquet et al., 1984). Thus the deformylase enzyme appears to be less stable than the formylase enzyme. Moreover, in modern bacteria the toxic activity of formylation is likely to be intrinsically more stable than deformylation – formylation creates a pool of fMet-$tRNA_i^{fMet}$ with a higher affinity for initiation factor 2 than Met-$tRNA_i^{fMet}$ (Steiner-Mosonyi et al., 2004). Consequently, even after any formylase activity has ceased, fMet-$tRNA_i^{fMet}$ will be preferentially used to initiate translation. In order to relieve the toxicity of formylation, deformylase activity must be more stable than the combined stability of formylase activity

and the fMet-tRNA$_i^{fMet}$ pool. Thus the *def-fmt* gene pair likely shows the differential toxin-antitoxin stability required to elicit PSK.

The central question of this part of my work was to determine if it was plausible that the *def-fmt* gene pair evolved as a toxin-antitoxin pair which could exhibit PSK activity when borne by horizontally mobile elements. To examine whether the *def-fmt* gene pair is able to elicit PSK, we introduced the *def-fmt* operon to our DF200 lines on a plasmid vector. We show that formylase and deformylase expression in DF200 does not show a phenotype consistent with complementation. Instead, the *def-fmt* operon shows a phenotype consistent with toxin-antitoxin activity, suggesting the spread of formylation through bacteria may have been a result of competition between mobile genetic elements mediated by PSK.

Methods

Strains and media

E. coli strain REL606 was obtained from T. Cooper (University of Houston, Texas). Strains DF000 and DF200 were developed during this work (Chapters 2 and 3, respectively). All media additives were purchased from Sigma-Aldrich unless otherwise specified. *E. coli* strains were grown at 37°C in Davis Minimal Broth (Difco) supplemented with 2mg/L thiamine and either 200mg/L dextrose (DM200) or 2000mg/L dextrose (DM2000), unless otherwise specified. When required, antibiotics were used at the following concentrations: Ampicillin, 100µg/mL; Streptomycin, 100µg/mL; Kanamycin, 20µg/mL; Chloramphenicol, 20µg/mL; Tetracycline 10µg/mL; Actinonin, 100µg/mL (Peptides International).

Plasmid construction

The *tetA* gene was amplified from plasmid RP4 (GenBank: BN000925.1) using the following primers:

TetA-F - 5′ – GGATCACTGTATTCGGCTGCAAG̊C̊TTGTCATGC – 3′ where dots above indicate mismatches used to generate *Hin*dIII restriction site (underlined), and

TetA-R - 5′ – CGTTCCAACCGÀAṪTCCTAGGGCGC – 3′ where dots above indicate mismatches used to generate EcoRI restriction site (underlined).

The *TetA* amplification product was cloned between the *Hin*dIII and *Eco*RI restriction sites of the mini-F plasmid pMF3 (Manis and Kline, 1977), replacing the *bla-cad′* cassette. A synthesized multiple cloning site (Integrated DNA Technologies) was cloned in to the *Bam*HI site of the pMF3-*TetA* plasmid. The resulting plasmid was confirmed by sequencing (Macrogen Korea) and named pF-0 (restriction map shown in Supplementary Figure 1). The *def-fmt* operon was amplified from the REL606 genome using the following primers:

def-180 - 5′ – GTACAAGCTGCTGATACTCATTAAACG – 3′

fmt+8 - 5′ – GTGGACTATCAGACCAGACGG – 3′

The resulting amplicon was cloned in to the pCR-Blunt vector (Invitrogen) following the manufacturer's instructions. The resulting plasmid was digested with *Bam*HI and *Not*I, and the *def-fmt* containing fragment cloned in to the corresponding restriction sites of pF-0. The resulting plasmid was confirmed by sequencing (Macrogen Korea) and named pF-DF.

Temperature sensitive plasmid construction

Plasmids pHSG415 and pTN9 were obtained from Ichizo Kobayashi (University of Tokyo, Japan) and renamed pTS and pTS::*PaeR7*, respectively. pTS::*def-fmt* was generated by digesting plasmid pJET::*def-fmt* with *Bgl*II and *Xho*I and cloning the

def-fmt containing fragment between the *Bam*HI and *Xho*I sites of pTS. Plasmids were confirmed by sequencing (Macrogen Korea).

RT-PCR

RNA was isolated from stationary phase cultures by the hot phenol method (Schmitt et al. 1990). Purified RNA was diluted to 200ng/µL and treated with TURBO DNase (Ambion) following the manufacturers guidelines. RT-PCR was carried out using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen) following the manufacturers guidelines. Primers used for RT-PCR were as follows:

*def*_internal_fwd: 5′ – CGAGACGATGTACGCAGAAGAAGG – 3′

*fmt*_internal_rev: 5′ – GAACTTCTGGTTTCGCCGTGCC – 3′

Primers used to detect *gstA* transcripts were:

*gstA*_fwd - 5′ – CTTTGCCGTTAACCCTAAGGG-3′

*gstA*_rev - 5″ – GCTGCAATGTGCTCTAACCC – 3′.

Northern blotting

RNA was isolated from stationary phase cultures under acidic conditions as described previously (Köhrer and RajBhandary, 2008). Aminoacylated tRNA in the total cellular RNA sample was deacylated by the addition of Tris-HCl, pH 9.5 to a final concentration of 0.1M followed by incubation at 37°C for 1 hour. Non-formylated aminoacyl-tRNA was deacylated by the addition of $CuSO_4$ and sodium acetate, pH 5.0 to a final concentration of 10mM and 0.1M, respectively, followed by incubation at 37°C for 30 minutes. Approximately 400ng of total RNA was separated by electrophoresis using 6.5% polyacrylamide (19:1 acrylamide:bisacrylamide), 8M urea, 0.1M sodium acetate pH 5.0 at 300V for 48

hours at 4°C. The initiator tRNA consistently migrated between the xylene cyanol and bromophenol blue loading dye bands. A 10cm long portion of the gel containing the tRNA was transferred to a Hybond-N+ membrane (GE Healthcare Life Sciences) in 4x TE buffer (40mM Tris-HCl, 4mM EDTA, pH 8.0) at 10V for 16 hours. RNA was fixed to the membrane by baking at 80°C for 2 hours. The membrane was washed twice with 6x SSC (1xSSC is 150mM NaCl, 15mM trisodium citrate, pH 7.0) followed by 4X SSC and 2x SSC.

$tRNA_i$ was probed using a 5' Digoxigenin labeled (NHS-ester) synthetic DNA probe (Integrated DNA Technologies) corresponding to bases 20-45 of the *E. coli* initiator tRNA (5′ – DIG-NHS-CTTCGGGTTATGAGCCCGACGAGCTA – 3′) and detected using Anti-Digoxigenin-AP Fab fragments (Roche) and CPD-*Star* (Roche) as per the manufacturer's instructions. Chemiluminescence was detected using a G:Box gel documentation system (Syngene).

Complementation test

Plasmids pF-0 and pF-DF were used to transform DF200. The resulting strains were grown to saturation in DM200 containing tetracycline. Cultures were diluted 1:100 in fresh media and the OD_{595} of these cultures was then monitored for 16 hours at 37°C (with shaking) using the FLUOstar OPTIMA. The growth rate was determined as the minimum doubling time taken over a 30-minute interval.

Determination of def-fmt gene order

Genome coordinates for the *def* or *fmt* gene(s) were collected from 940 bacterial genome sequences. The coordinates were examined to determine the order and adjacency of the *def* and *fmt* genes. Genes were considered to be in a putative operon arrangement if they are encoded on the same strand and intergenic space is less than 100bp.

Plasmid stability

The stability of plasmids in culture was measured using a repeated batch culture method (Nordström, 1993). DF200 strains containing plasmids pTS, pTS::*PaeR7,* or pTS::*def-fmt* were grown in DM2000 containing Streptomycin, Ampicillin and Chloramphenicol. Cultures were plated on solid media with and without antibiotics to determine proportion of plasmid containing cells. Cultures were then, pelleted, washed twice with PBS and diluted 10^6-fold in DM2000 containing Streptomycin and grown for 24 hours at 30°C. This process was repeated three times with plasmid occupancy measured by plating at each passage.

Post-segregational killing

Strain DF200 carrying plasmids pTS, pTS::DF or pTS::R7 was used to measure post-segregation killing through temperature-induced plasmid loss as described previously (Naito et al., 1995). Briefly, cultures were grown to an OD_{600} of 0.3 at 30°C before diluting 1:20 in fresh media and growing at 42°C. In order to maintain exponential growth, cultures were diluted 1:20 whenever the OD_{600} reached 0.3. At 1-hour time intervals, serial dilutions of each culture were plated on DM2000 agar (to measure total viable cells) and DM2000 containing Chloramphenicol and Ampicillin (to measure plasmid-containing cells). Plates were grown for 72 hours at 30°C before colonies were enumerated.

Results

The def-fmt gene pair resembles a toxin-antitoxin operon

As the *def-fmt* gene pair satisfies the toxin/antitoxin and stability requirements to elicit PSK, we asked whether the *def-fmt* gene pair is arranged in a similar operonic structure to modern toxin-antitoxin gene pairs. Many TAs are arranged in a bicistronic operon with the antitoxin gene encoded upstream of the toxin. It is thought that this arrangement prevents potentially lethal expression of the toxin before the antitoxin (Pandey and Gerdes, 2005). In *E. coli,* the *def* gene of the *def-fmt* gene pair is encoded upstream of the *fmt* gene in a biscistronic operon (Meinnel and Blanquet, 1993). To examine whether this *def-fmt* arrangement was preserved in other bacterial genomes, we conducted a simple comparative genomic study. We surveyed 940 bacterial genomes for the relative positions of the *def* and *fmt* genes. The *def* and *fmt* genes were considered to be in a putative operon arrangement if they encoded in the same orientation with an intergenic space less than 100bp. This operonic *def-fmt* arrangement was found to be preserved in half of genomes analysed (Figure 1). In contrast, no cases were observed where the *fmt* gene was encoded upstream of the *def* gene within a putative operon. In genomes where the *def* and *fmt* genes are encoded greater than 100bp apart, this apparent requirement for gene order is lost. Although this arrangement is not essential for TA activity, the *def-fmt* gene order is clearly preserved in many bacterial genomes.

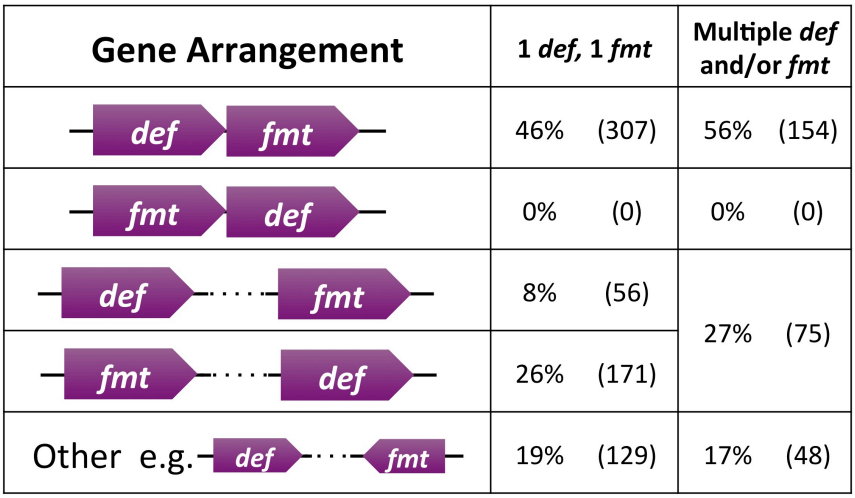


Figure 1 | A *def-fmt* operon arrangement is preserved in many bacterial genomes. The proximity of *def* and *fmt* genes was analysed in 940 bacterial genomes. Where *def* and *fmt* are present in single copies, a *def-fmt* putative operon arrangement is observed in almost half of genomes, with no *fmt-def* arrangements observed. This arrangement is not conserved when genes are non-adjacent. In genomes where multiple *def* and/or *fmt* genes are present, a *def-fmt* operon is preserved in 56% of cases, with no observed *fmt-def* arrangements. The presence of multiple gene copies in circular bacterial genomes makes gene order meaningless if genes are not adjacent, thus the non-adjacent arrangements are combined.

To test if the *def-fmt* operon can elicit a PSK phenotype in DF200, we tested the *def-fmt* operon for two traits commonly conferred by TA gene pairs: increased plasmid stability in monoculture and cell death upon plasmid loss.

Plasmid stability

Plasmids encoding TA gene pairs appear to exhibit increased stability in a culture through increased occupancy of cells. This phenotype is not a result of increased plasmid stability, but rather the killing of cells that do not harbor the plasmid following cell division (Figure 2).

To test whether the *def-fmt* operon can increase the apparent stability of a plasmid in our DF200 line, we measured the occupancy rates of plasmids with and without the *def-fmt* operon during serial culturing. We cultured DF200 containing either pTS (vector control), pTS::*PaeR7* (encoding the *PaeR7* restriction-modification system – a toxin-antitoxin system known to elicit PSK (Naito et al., 1995)) or pTS::*def-fmt* (encoding the *def-fmt* operon) for 3 serial passages of culturing without antibiotic selection. Under these conditions, a control plasmid without a TA gene pair (pTS) was quickly lost from the culture (Figure 3a), detectable in only 1-3% of the culture population occupancy after 3 passages. The presence of a known toxin-antitoxin gene pair (pTS::*PaeR7*) resulted in a greater apparent stability (Figure 3b), increasing plasmid occupancy to 25-80% after 3 passages. The presence of the *def-fmt* operon increased the occupancy of the plasmid (pTS::*def-fmt*) to 15-60% after 3 passages (Figure 3c). This plasmid-stabilising nature of the *def-fmt* operon was not observed in REL606 (Figure 3d) where the *def-fmt* gene pair is encoded on the chromosome. This is consistent with the ability of chromosomally-encoded TAs to immunize the host from PSK elicited by plasmid-encoded TAs (Cooper and Heinemann, 2000; 2005).

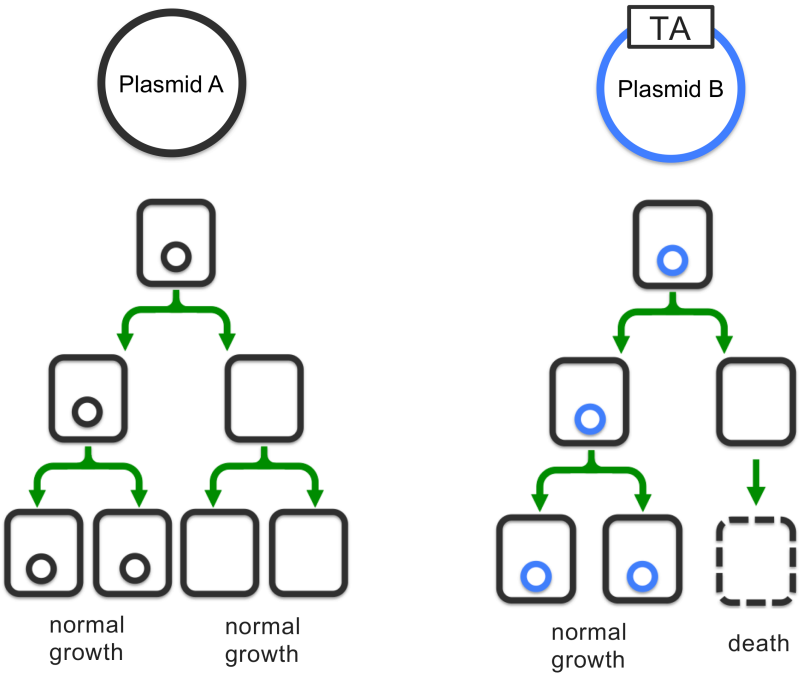


Figure 2 | TA-encoding plasmids show increased occupancy in cultures. Plasmids can be lost due to missegregation during cell division, resulting in a portion of the population becoming plasmid-free. If plasmids do not encode a TA system (Plasmid A), both plasmid-containing and plasmid-free cells are able to reproduce, decreasing the proportion of the population infected by the plasmid. In contrast, when a plasmid encodes a TA gene pair that can elicit PSK (Plasmid B), the plasmid may missegregate during cell division. However, in this case plasmid loss causes cell death. As a result, the plasmid remains at high occupancy in the culture, displaying an apparent increase in plasmid stability.

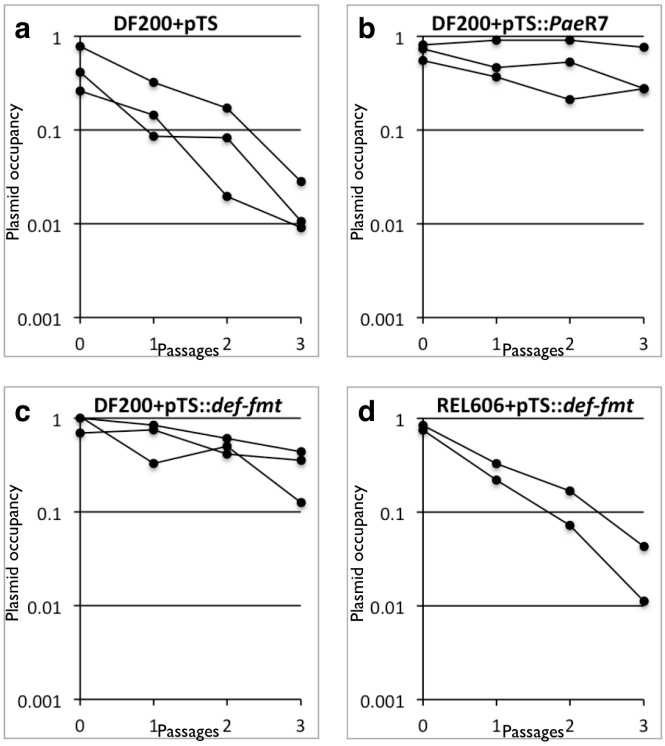


Figure 3 | The *def-fmt* gene pair increases the stability of plasmids in DF200. a) DF200 containing pTS was serially cultured with daily 10^6 dilutions in fresh media at 30°C without antibiotics. Plasmid occupancy was measured by plating cultures in triplicate on media with and without selection for plasmid-encoded antibiotic markers. Plasmid-free cells quickly accumulated in the DF200+pTS culture. **b)** Plasmid pTS::*PaeR7* is maintained in cultures at high occupancy. **c)** The *def-fmt* gene pair also stabilizes plasmids in DF200. **d)** pTS::*def-fmt* does not show heightened stability in a REL606 population.

Post-segregational killing

The ability of a *def-fmt* gene pair to stabilize a plasmid in DF200 is consistent with a PSK phenotype. However, small differences in the growth rate of cells with and without *def-fmt* encoding plasmids may result in an apparent increase in stability without requiring PSK. Such a difference in growth rate could be due to the potential for the *def-fmt* gene pair to provide an additional fitness benefit even to the evolved Δ(*def-fmt*) strain. Therefore, we next examined whether a *def-fmt*-encoding plasmid is able to elicit PSK when the production of plasmid-free segregants is forced by halting plasmid replication.

To test whether a *def-fmt* encoding plasmid can elicit a PSK phenotype in DF200, we used a sensitive vector system with a temperature-dependent replication trait to artificially force the formation of plasmid-free cells in a culture of plasmid-containing cells. The temperature sensitive vector, pTS, is unable to replicate at 42°C. Consequently, plasmid replication fails to keep pace with cell division, resulting in the formation of plasmid-free daughter cells. Therefore, each cell division on average produces one plasmid-free cell and one plasmid-containing cell. If the vector encodes a TA gene pair that can elicit PSK, the plasmid-free daughter cell will be killed (Figure 4).

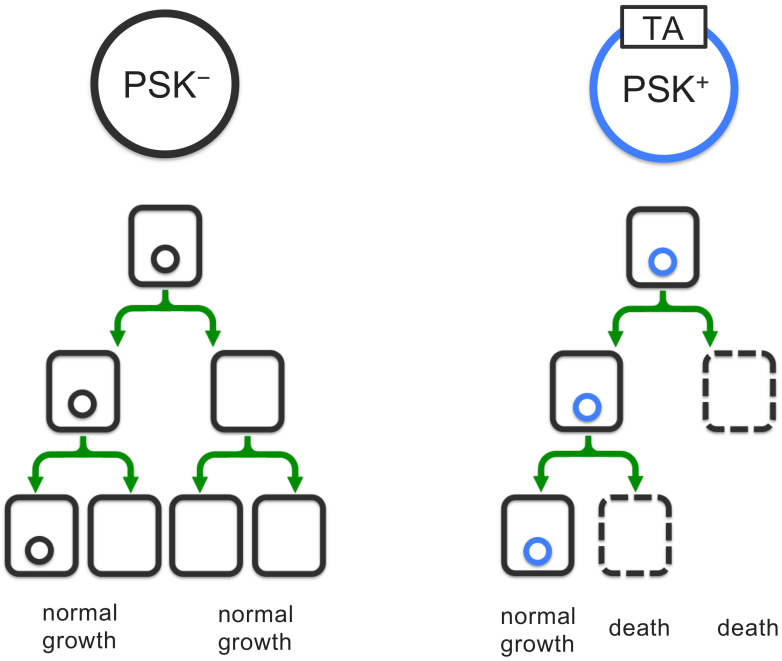


Figure 4 | Temperature sensitive plasmids induce PSK. When a culture of bacteria is grown at a temperature non-permissive for plasmid replication, the temperature-sensitive plasmid fails to replicate at the same pace as cell division. If this plasmid contains a TA gene pair, the resulting missegregation of the plasmid can cause cell death, thus demonstrating the diagnostic for PSK activity.

We grew DF200 cells containing plasmid pTS, pTS::*def-fmt* or pTS::*PaeR7* at 30°C (permissive for plasmid replication) to early log phase with antibiotic selection for plasmid-encoded resistance markers. At this stage of the experiment, the antibiotic serves as an external toxin with the resistance gene the internal antitoxin. The antibiotic ensures that all viable daughter cells arising prior to the temperature shift contain the plasmid.

Cultures were then diluted in fresh media without antibiotics and grown at 42°C (prohibitive to plasmid replication). Cultures were diluted regularly to keep them in logarithmic growth, and samples taken at regular intervals to measure culture growth and plasmid occupancy.

When cells harbour a plasmid without a TA gene pair, we expect plasmid replication to cease at 42°C and plasmid-free cells to dominate the culture, similar to that described in Figure 4. Indeed, this is what we observed with DF200+pTS following transfer from 30°C to 42°C (Figure 5a). The relative optical density of the culture continued to increase exponentially throughout the 8 hours of culturing at 42°C. The total viable cells also increased exponentially throughout culturing, indicating that the increase in optical density was due to the formation of viable cells. The number of cells containing the pTS plasmid did not increase during the 8 hours at 42°C, indicating that plasmid replication had ceased, despite ongoing cell replication (Figure 5a).

When cells harbor a plasmid capable of eliciting PSK, we expect plasmid-free cells to be killed by the action of the toxin at cell division, resulting in the production of non-viable cells (Figure 4). This trend is observed with DF200+pTS::*PaeR7* where the number of viable cells in the culture remained static for approximately 4 hours following the shift to 42°C (Figure 5b). Despite this apparent stasis of culture growth, the relative optical density of the culture continued to increase throughout the experiment. However, as the increase in optical density is produced by the

accumulation of non-replicating cells, the increase is not exponential as with pTS. The number of plasmid-containing cells remains constant for the 8 hours of the experiment, showing an inability of the plasmid to replicate at 42°C. After approximately 4 hours of culture, the number of viable cells began to increase. This has been observed previously (Naito et al., 1995) and is likely due to the appearance of individuals who have escaped death by PSK (Cooper and Heinemann, 2000). We expect the frequency of cells escaping PSK to be very small under standard culturing conditions, requiring plasmid missegregation followed by escape from the toxic effects. However, in this experimental setup the frequency of plasmid missegregation is expected to be near 1.0, allowing observation of rare survivors of PSK.

When DF200 cells harbor the *def-fmt* encoding plasmid pTS::*def-fmt*, the culture demonstrates the same characteristics as our pTS::*PaeR7* control culture (Figure 5c). That is, the *def-fmt* operon is able to elicit a PSK phenotype indistinguishable from the *PaeR7* restriction-modification system. Following loss of the *def-fmt*-encoding plasmid from DF200, we expect that continued formylase activity and the formylated Met-$tRNA_i^{fMet}$ pool would result in translation initiation using fMet-$tRNA_i^{fMet}$. However, in the absence of deformylase expression, residual deformylase activity would quickly be lost, resulting in the accumulation of formylated peptides. Without restoration of deformylase activity, this retention of the formyl-group prevents the essential *N*-terminal processing of peptides, resulting in cell death.

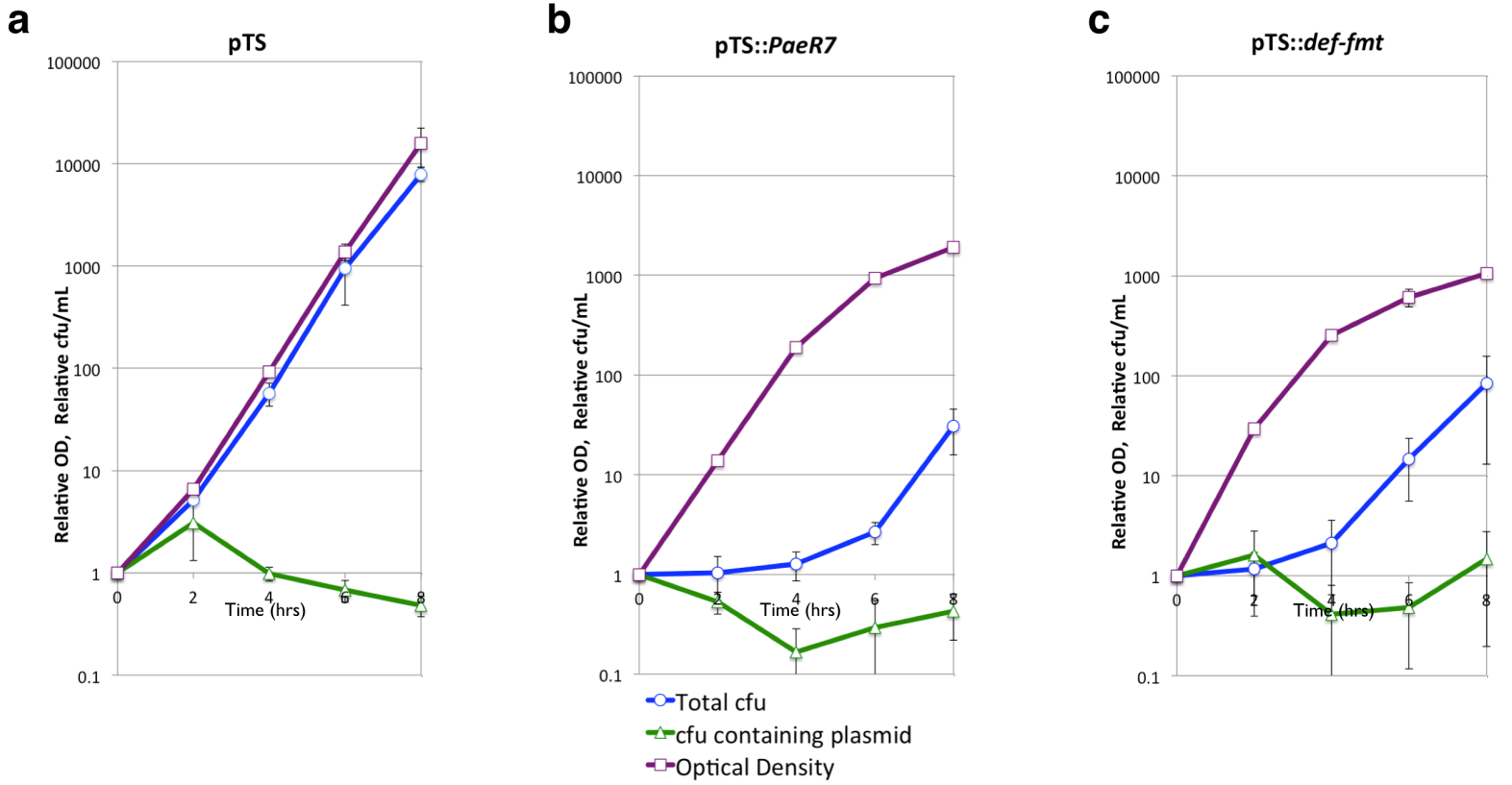
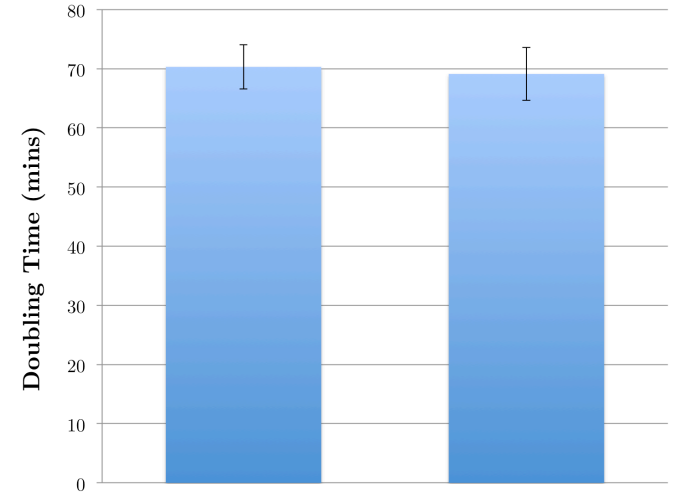


Figure 5 | A plasmid-borne *def-fmt* operon demonstrates in culture activities diagnostic of PSK activity. A DF200 strain containing pTS, pTS::*PaeR7* or pTS::*def-fmt* was cultured to an OD of 0.3 at 30°C with antibiotic selection for plasmid markers, then transferred to media without selection and grown at 42°C to prevent plasmid replication. Time points (x-axis) are shown as hours following transfer to 42°C. Cultures were maintained in exponential growth by diluting in fresh media when the OD reached 0.3. **a)** DF200+pTS showed increased exponentially in total cfu and optical density, while plasmid containing cells (showing chloramphenicol and ampicillin resistance) remained almost unchanged. **b)** DF200+pTS::*PaeR7* increased very little in viable cell count for 4 hours after transfer to 42°C, despite continued cell division (OD600). This phenotype is consistent with PSK. **c)** DF200+pTS::*def-fmt* demonstrated growth characteristics similar to DF200+pTS::*PaeR7*, suggesting the *def-fmt* gene pair elicits PSK in DF200.

Expression of def-fmt does not alter the growth rate of DF200

To further establish that the relative increase in the accumulation of the *def-fmt* plasmid was not due to latent complementation conferring a fitness advantage, growth rates were monitored. The ability of the *def-fmt* gene pair to increase the apparent stability of a plasmid in DF200 could be an artifact caused by differences in the growth rate of plasmid-containing and plasmid-free cells. If DF200 cells containing *def-fmt*-encoding plasmids grew significantly faster than cells without a *def-fmt*-encoding plasmid, plasmid-containing cells would accumulate faster in the population. This would be indistinguishable from a PSK-induced stability phenotype. Additionally, all 11 of our DF200 lines had mutations in the C2 domain of initiation factor 2 (Chapter 3, Figure 7). These mutations were expected to change the binding affinity of IF-2 for Met-$tRNA_i^{fMet}$ such that translation initiation is more efficient when fMet-$tRNA_i^{fMet}$ is absent. Conversely, such mutations may decrease the efficiency of translation in the presence of fMet-$tRNA_i^{fMet}$, slowing the growth of cells.

To examine whether formylation has an effect on the growth rate of DF200, we measured the minimum doubling time of DF200 with and without a plasmid-borne *def-fmt* operon. The presence of the *def-fmt* gene pair had no effect on the maximum growth rate of DF200 (Figure 6). This result also suggests that, unlike DF000, our DF200 line no longer benefits from expression of the *def-fmt* operon from a plasmid vector.



Strain	DF200+pF-0	DF200+pF-DF
Doubling Time	70±4	69±4

Figure 6 | Expression of plasmid-encoded *def-fmt* does not change the growth rate of DF200. DF200 containing either pF-0 or pF-DF was grown in DM2000 media supplemented with tetracycline. The doubling time was measured as the minimum OD595 doubling time taken over a 30-minute interval. The doubling times of DF200+pF-0 and DF200+pF-DF were 70±4 and 69±4 minutes, respectively. Samples shown as $\overline{X}$±SE, n=5.

Plasmid-encoded def-fmt is expressed and active in DF200

Although the *def-fmt* gene pair is able to elicit a PSK phenotype in DF200, the indistinguishable growth rates of strains with and without the *def-fmt* gene pair suggest that formylation may not be functioning as expected in DF200. To confirm expression of the *def-fmt* operon from pF-DF, we used RT-PCR to identify bicistronic *def-fmt* transcripts. *def-fmt* transcripts were detected in DF200 carrying pF-DF, but not in a strain carrying pF-0, or plasmid-free cells using RT-PCR (Figure 7). Thus, the *def-fmt* operon was transcribed from the pF-DF plasmid.

Our DF200 lines had no mutations in the genes encoding $tRNA_i^{fMet}$ (*metZ, metW, metV* or *metY*), thus the initiator tRNA in DF200 encodes the features required for formylation (see Chapter 1, Figure 5). As a result, expression of the formylase enzyme in DF200 should be functional. Northern blotting was used to confirm the absence of formylated tRNA. Total RNA was deacylated by incubation at pH 9.5 (resulting in complete deacylation of aminoacyl-tRNA e.g. both fMet-$tRNA_i^{fMet}$ and Met-$tRNA_i^{fMet}$), or in 10mM $CuSO_4$, pH 5.3 (resulting in deacylation of non-formylated aminoacyl-tRNA e.g. Met-$tRNA_i^{fMet}$). REL606 $tRNA_i^{fMet}$ was resistant to deacylation by $CuSO_4$, but sensitive to deacylation at pH 9.5 (Figure 8a), indicating that the *N*-terminus of the REL606 Met-$tRNA_i^{fMet}$ was blocked by formylation. In contrast, $tRNA_i^{fMet}$ isolated from DF200 was sensitive to deacylation by Cu^{2+} (Figure 8b), indicating the Met-$tRNA_i^{fMet}$ had a free *N*-terminus i.e. no formylation. Expression of *def-fmt* from pF-DF in DF200 restored the resistance of $tRNA_i^{fMet}$ to deacylation by Cu^{2+}, and resulted in a $tRNA_i^{fMet}$ banding pattern identical to the REL606 (Figure 8c). This indicated that plasmid-encoded formylase was indeed active in DF200.

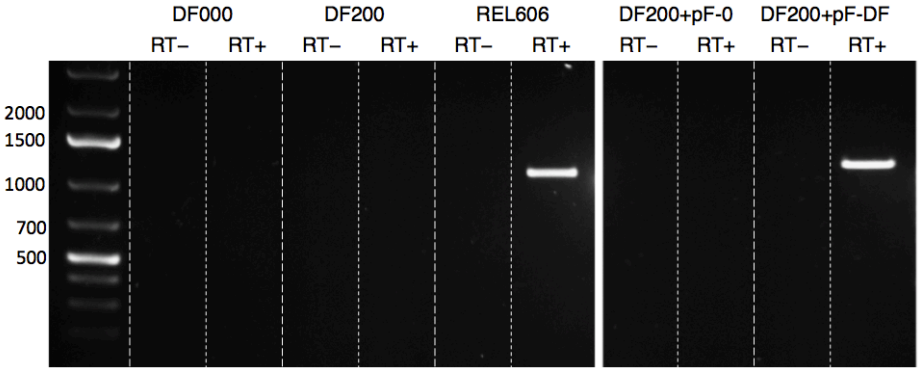


Figure 7 | Plasmid-encoded *def-fmt* was expressed in DF200. Total RNA was extracted from early stationary phase cultures of DF000 and REL606 and used as a template for RT-PCR. RT- lanes were processed without a reverse transcriptase reaction in order to detect the presence of genomic DNA. RT+ lanes were processed as per the manufacturer's guidelines using primers corresponding to REL606 *def-fmt* operon (expected size 1006bp). Lane 1, GeneRuler 1kb Plus (Thermo Scientific) with size standards indicated; lanes 2-7, RT-PCR products for DF000 and REL606 with RT- controls indicated; lanes 8-11, RT-PCR products for DF200 containing pF-0 or pF-DF with RT- controls indicated.

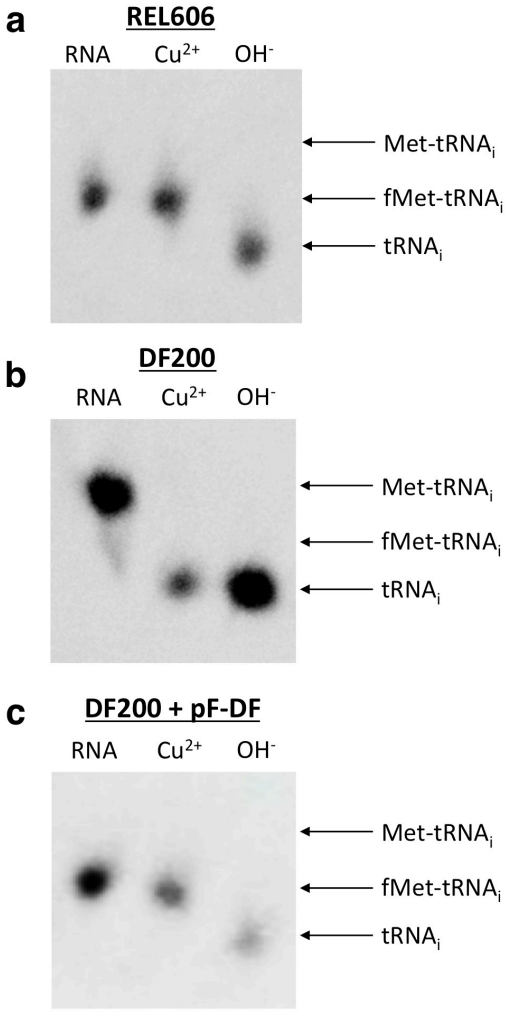
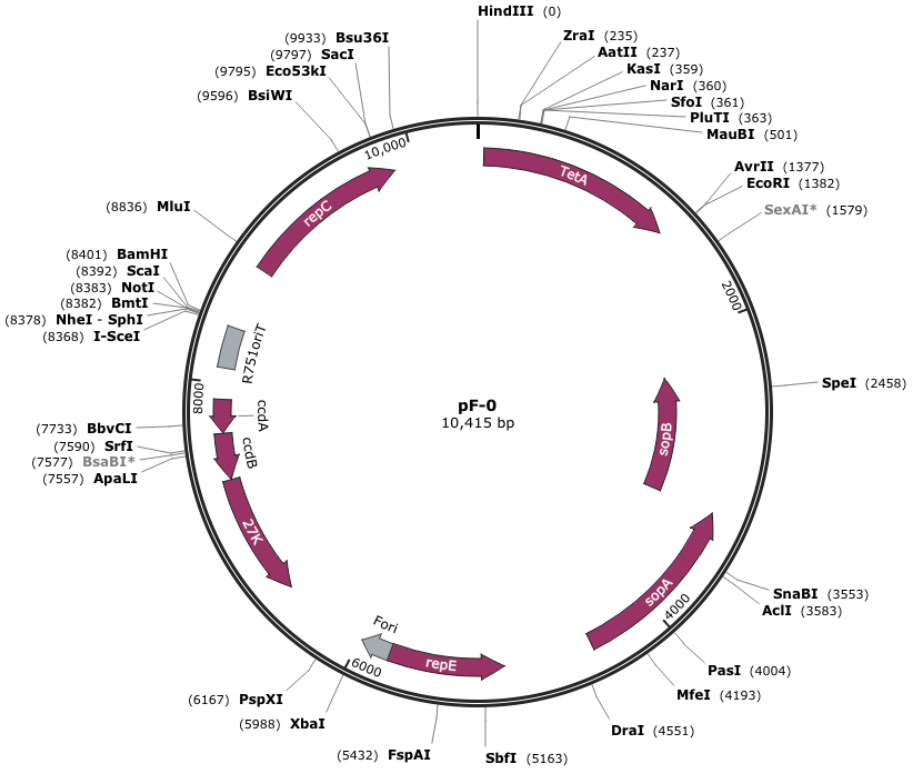


Figure 8 | Plasmid-encoded *def-fmt* restores formylation in DF200. Total RNA from REL606, DF200 or DF200 containing plasmid pF-DF was isolated under acidic conditions. tRNA was either left untreated (RNA) or deacylated by incubation at pH 5 with 10mM $CuSO_4$ (Cu^{2+} condition) or pH 9 (OH^- condition) and examined by urea-PAGE followed by northern blotting for $tRNA_i^{fMet}$. **a)** $tRNA_i^{fMet}$ from REL606 shows no change in electrophoretic mobility following treatment with $CuSO_4$ indicating Met-$tRNA_i^{fMet}$ is formylated. Incubation of RNA at pH 9 results in a change in mobility of the $tRNA_i^{fMet}$ consistent with deacylation. **b)** $tRNA_i^{fMet}$ from DF200 shows a change in electrophoretic mobility following treatment with $CuSO_4$. This change in mobility is identical to incubation of RNA at pH 9, and is consistent with deacylation, indicating Met-$tRNA_i^{fMet}$ is not formylated. **c)** $tRNA_i^{fMet}$ isolated from DF200 expressing *def-fmt* from pF-DF shows resistance to deacylation by $CuSO_4$ similar to REL606. This suggests Met-$tRNA_i^{fMet}$ is formylated in this strain.

Discussion

Formylation is a ubiquitous process throughout the bacterial domain (Yuan et al., 2001). Nevertheless, we have shown previously that formylation is not required for wild-type growth rates in *E. coli* (see Chapter 3). This apparent dispensability makes it difficult to reconcile formylation with a fixed and adaptive functional role. Considering this, we asked how formylation could have emerged and then spread to fixation in bacteria and be maintained over evolutionary time frames. We show that the *def-fmt* gene pair resembles modern toxin-antitoxin systems that confer PSK traits on mobile genetic elements. Additionally, reintroduction of the *def-fmt* operon to our evolved DF200 strain produced a phenotype that was not consistent with complementation. Rather, the *def-fmt* gene pair is able to elicit a PSK phenotype. These observations support the hypothesis that the *def-fmt* gene pair emerged in its modern form because of the fitness benefit to genetic elements that replicate infectiously, such as conjugative plasmids, consistent with the competition model (Heinemann, 1998).

Previous experiments have shown that TA gene pairs that elicit PSK have a propensity to propagate and spread through competition between horizontally mobile elements (Cooper and Heinemann, 2000; 2005; Cooper et al., 2010). This mechanism of evolution can be considered non-adaptive relative to the host because it does not require the TA gene pair to confer a selective benefit to host cells. The finding that the *def-fmt* gene pair can elicit PSK suggests that formylation could have spread through a non-formylating bacterial population by competition between horizontally mobile elements. This mechanism of evolution is not reliant on formylation conferring a selectively advantageous function to bacteria. Evolution by PSK may help to explain the contradiction between the ubiquity of formylation and its dispensability.



Supplementary Figure 1 | Plasmid map of pF-0

References

Adams, J.M. (1968). On the release of the formyl group from nascent protein. J. Mol. Biol. *33*, 571–589.

Blanquet, S., Dessen, P., and Kahn, D. (1984). Properties and specificity of methionyl-tRNAfMet formyltransferase from Escherichia coli. Meth. Enzymol. *106*, 141–152.

Cooper, T.F., and Heinemann, J.A. (2000). Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. PNAS *97*, 12643–12648.

Cooper, T.F., and Heinemann, J.A. (2005). Selection for plasmid post-segregational killing depends on multiple infection: evidence for the selection of more virulent parasites through parasite-level competition. Proc. R. Soc. B *272*, 403–410.

Cooper, T.F., Paixão, T., and Heinemann, J.A. (2010). Within-host competition selects for plasmid-encoded toxin-antitoxin systems. Proc. R. Soc. B *277*, 3149–3155.

Engelberg-Kulka, H., Sat, B., Reches, M., Amitai, S., and Hazan, R. (2004). Bacterial programmed cell death systems as targets for antibiotics. Trends in Microbiology *12*, 66–71.

Gerdes, K., Rasmussen, P.B., and Molin, S. (1986). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. PNAS *83*, 3116–3120.

Gerdes, K., Christensen, S.K., and bner-Olesen, A.L.O. (2005). Prokaryotic toxin-antitoxin stress response loci. Nature Reviews Microbiology *3*, 371–382.

Hayes, F. (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science *301*, 1496–1499.

Hazan, R., and Engelberg-Kulka, H. (2004). Escherichia coli mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol. Genet. Genomics *272*, 227–234.

Heinemann, J.A. (1998). Looking sideways at the evolution of replicons. In Horizontal Gene Transfer, M. Syvanen, and C.I. Kado, eds. (London: Academic Press), pp. 11–24.

Köhrer, C., and RajBhandary, U.L. (2008). The many applications of acid urea

polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases. Methods *44*, 129–138.

Kusano, K., Naito, T., Handa, N., and Kobayashi, I. (1995). Restriction-modification systems as genomic parasites in competition for specific sequences. PNAS *92*, 11095–11099.

Livingston, D.M., and Leder, P. (1969). Deformylation and protein biosynthesis. Biochemistry *8*, 435–443.

Manis, J.J., and Kline, B.C. (1977). Restriction endonuclease mapping and mutagenesis of the F sex factor replication region. Molec. Gen. Genet. *152*, 175–182.

Mazel, D., Pochet, S., and Marliere, P. (1994). Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO J. *13*, 914–923.

Meinnel, T., and Blanquet, S. (1993). Evidence that peptide deformylase and methionyl-tRNA(fMet) formyltransferase are encoded within the same operon in Escherichia coli. Journal of Bacteriology.

Naito, T., Kusano, K., and Kobayashi, I. (1995). Selfish behavior of restriction-modification systems. Science *267*, 897–899.

Nordström, K. (1993). Plasmids: a practical approach. K.G. Hardy, ed. (Oxford University Press), p. 280.

Ogura, T., and Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. PNAS *80*, 4784–4788.

Pandey, D.P., and Gerdes, K. (2005). Toxin–antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res. *33*, 966–976.

Pecota, D.C., and Wood, T.K. (1996). Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. J. Bacteriol. *178*, 2044–2050.

Rajagopalan, P., Datta, A., and Pei, D. (1997). Purification, Characterization, and Inhibition of Peptide Deformylase from Escherichia coli† - Biochemistry (ACS Publications). Biochemistry.

Rankin, D.J., Turner, L.A., Heinemann, J.A., and Brown, S.P. (2012). The coevolution of toxin and antitoxin genes drives the dynamics of bacterial addiction complexes and intragenomic conflict. Proc. R. Soc. B *279*, 3706–3715.

Steiner-Mosonyi, M., Creuzenet, C., Keates, R.A.B., Strub, B.R., and Mangroo, D. (2004). The Pseudomonas aeruginosa initiation factor IF-2 is responsible for formylation-independent protein initiation in P. aeruginosa. J. Biol. Chem. *279*, 52262–52269.

Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe Magnus, D.A. (2007). Chromosomal toxin–antitoxin loci can diminish large-scale genome reductions in the absence of selection. Mol. Microbiol. *63*, 1588–1605.

Van Melderen, L. (2010). Toxin-antitoxin systems: why so many, what for? Curr. Opin. Microbiol. *13*, 781–785.

Yuan, Z., Trias, J., and White, R.J. (2001). Deformylase as a novel antibacterial target. Drug Discovery Today *6*, 954–961.

Chapter 5
def-fmt cannot elicit PSK in *S. cerevisiae*

All experimental work and subsequent analyses outlined in this chapter were performed by Ryan Catchpole under the mentorship of Prof. Jack Heinemann and Assoc. Prof. Anthony Poole.

Introduction

Translation is initiated in eukaryotes using a methionylated initiator tRNA (Met-$tRNA_i^{Met}$) (Smith and Marcker, 1970). This is in contrast to bacteria, where translation is initiated using a formylated methionyl-tRNA (fMet-$tRNA_i^{fMet}$) (Marcker and Sanger, 1964). The formylation process that generates fMet-$tRNA_i^{fMet}$ is preserved in mitochondria and plastids, the bacteria-derived organelles of eukaryotes (Giglione and Meinnel, 2001; Smith and Marcker, 1968). This conservation indicates that formylation was likely present in the ancestors of both mitochondria and plastids. Genes encoding the formylase and deformylase enzymes are no longer present in the mitochondrial or chloroplast genomes, having been transferred to the nuclear genome (Giglione, 2000). Despite this transfer, the mitochondrial translation machinery has retained a bacteria-like affinity for formylated initiator tRNA. That is, mitochondrial initiation factor-2 (IF2) has a higher affinity for fMet-$tRNA^{Met}$ than Met-$tRNA^{Met}$ (Garofalo et al., 2003; Liao and Spremulli, 1991), similar to bacterial IF2. In contrast, eukaryotic cytoplasmic IF2 has a high affinity for non-formylated Met-$tRNA_i^{Met}$ (Naveau et al., 2010).

Interestingly, the mitochondrial genome of animals encodes a single $tRNA^{Met}$, which is used in both translation initiation and elongation. This $tRNA^{Met}$ is methionylated and formylated for use in translation initiation, and used in a methionylated but non-formylated state for elongation (Takeuchi et al., 1998). Formylation of initiator tRNA is achieved through targeting of the nuclear-encoded formylase enzyme to the mitochondria where it recognizes and formylates Met-$tRNA^{Met}$. Similar to bacteria, translation initiation with fMet-$tRNA^{Met}$ produces *N*-formylated peptides which require deformylation. This is achieved through targeting of the nuclear-encoded deformylase enzyme to the mitochondria. This targeting of formylase and deformylase enzymes to both the mitochondria and plastids is directed by signal peptides at the *N*-terminus of these proteins (Carrie and Small, 2013).

The nuclear-encoded formylase and deformylase genes of eukaryotes could have transferred to the nuclear genome through horizontal gene transfer (HGT) (from free living bacteria) or endosymbiont gene transfer (EGT) (from the ancestral mitochondria or chloroplasts). HGT between bacteria and eukaryotes has been identified as the source of numerous eukaryotic genes (Andersson, 2005). Many vectors of gene transfer between bacteria can also transfer genes from bacteria to eukaryotes (Heinemann and Kurenbach, 2009; Heinemann and Sprague, 1989; Ream, 1989). Additionally, EGT is an ongoing process with even extant mitochondrial genes being transferred to the nucleus frequently (Huang et al., 2003; Stegemann et al., 2003). As such, it is reasonable to expect that *def-fmt* gene pairs will have transferred from both bacteria and mitochondria/plastids to the nucleus frequently during eukaryote evolution. We showed previously (see Chapter 4) that the *def-fmt* gene pair elicits an addictive phenotype through post-segregational killing in bacteria devoid of formylation activity. As eukaryotes do not use formylation in translation, we would expect transfer of the *def-fmt* gene pair to the eukaryotic nucleus to result in addiction of cytoplasmic translation to the formylase-deformylase cycle. However, cytoplasmic translation in modern

eukaryotes proceeds without a formylated Met-tRNA. To determine how eukaryotes might have escaped this addictive phenotype, we expressed the *Escherichia coli* formylase and deformylase genes in *Saccharomyces cerevisiae*. We show that, in contrast to bacteria, expression of formylase in the absence of deformylase is not lethal in *S. cerevisiae*. Consequently, a *def-fmt* gene pair can be lost from yeast cytoplasmic translation without eliciting a post-segregational killing phenotype.

Methods

Strains and Media

S. cerevisiae strain W303α was obtained from N. Hollingsworth (Stony Brook University, NY, USA). Plasmid pBEVY-GA was obtained from C. Miller (Tulane University, LA, USA). Yeast strains were grown in SD-ade (0.67% Difco Yeast Nitrogen Base without Amino Acids, 2% gextrose, 30mg/L L-leucine, 20mg/L L-tryptophan, 20mg/L L-histidine and 20mg/L uracil) or SGal-ade (SD-ade media where dextrose is replaced with 2% galactose), at 30°C. Solid media was prepared by adding 1.5% (w/v) bacteriological agar (Oxoid) to SD-ade, SGal-ade, YPD (1% Yeast Extract, 2% Peptone, 2% Glucose) or YPGal (1% Yeast Extract, 2% Peptone, 2% Galactose). Where indicated, Actinonin was added to a final concentration of 200µg/mL.

pBEVY-GA expression plasmids

The *def* and *fmt* genes were amplified by PCR from REL606 genomic DNA. Primers used for amplification were designed to introduce restriction sites for subsequent cloning, as well as Kozak sequences to facilitate expression in a eukaryotic host. As the native start codon of the *fmt* gene in *E. coli* is GTG, primers were used to modify the start codon to ATG to allow expression in yeast. The primers used were as follows:

*def*_fwd: 5′ – CAGTGGTACC**GCCACC**ATGTCAGTTTTGCAAGTGTTACATATTCC – 3′

where underline indicates *Kpn*I recognition site, bold indicates Kozak sequence and waved underline indicates identity to nucleotides 1-29 of *def.*

*def*_rev: 5′ – CATGGAATTCGTTAGTTCTTATCCTTAAGCCCG – 3′

where underline indicates *EcoR*I recognition site, waved underline indicates identity to nucleotides 502-510 of *def* and 1-15 downstream of *def.*

*fmt*_fwd: 5′ –GTCAGGATCC**GCCACC**ATGTCAGAATCACTACGTATTATTTTTGCGG– 3′

where underline indicates *BamH*I recognition site, bold indicates Kozak sequence,

dot below indicates G→A change of start codon and waved underline indicates identity to nucleotides 2-31 of *fmt.*

*fmt*_rev: 5′ – CATGGTCGACGGGCTTAGAAGAGTGGACTATC – 3′

where underline indicates *Sal*I recognition site, waved underline indicates identity to nucleotides 947-948 of *fmt* and 1-20 downstream of of *fmt.*

The *def* amplicon was cloned between the *Kpn*I and *EcoR*I restriction enzyme sites of pBEVY-GA, downstream of the GAL1 promoter. The resulting plasmid was named pBEVY::*def*. The *fmt* amplicon was cloned between the *BamH*I and *Sal*I restriction enzyme sites of pBEVY-GA, downstream of the GAL10 promoter. The resulting plasmid was named pBEVY::*fmt.* Additionally, a plasmid was generated containing both the *def* and *fmt* genes. This plasmid was named pBEVY::*def,fmt.* Constructs were confirmed by sequencing (Macrogen Korea). Plasmid maps are provided in Supplementary Figure 1. The plasmids were used to transform yeast strain W303α to adenine prototrophy by electroporation as described previously (Becker and Guarente, 1991).

RT-PCR

RT-PCR was used to confirm transcription of *def* and/or *fmt* from plasmids. W303α strains harboring pBEVY-GA derived plasmids were grown to stationary phase in SGal-ade. Total RNA was isolated from cultures by the hot acid-phenol method (Schmitt et al., 1990). Purified RNA was diluted to 200ng/uL and treated with TURBO DNase (Ambion) following the manufacturers guidelines. RT-PCR was carried out using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen) following the manufacturers guidelines. Primers were as follows:

*def*_internal_fwd: 5′ – CGAGACGATGTACGCAGAAGAAGG – 3′

*def*_internal_rev: 5′ – CAGTTTGCCGACCAGGTGATCC – 3′

*fmt*_internal_fwd: 5′ – CCCGGTTAAAGTTCTGGCTGAGG – 3′

*fmt*_internal_rev: 5′ – GAACTTCTGGTTTCGCCGTGCC – 3′

*ADE2*_internal_fwd: 5′ – CCATACCTGGCAAGTGAGCAGC – 3′

*ADE2*_internal_rev: 5′ – CTCCACCATTACAACGAACGCC – 3′

Growth rates

In order to measure the growth rate of yeast expressing *E. coli* formylase and deformylase enzymes, W303α strains harboring either pBEVY-GA, pBEVY-GA::*def*, pBEVY-GA::*fmt*, or pBEVY-GA::*def,fmt* were grown in SD-ade media for 48 hours to stationary phase. Cultures were then pelleted, washed twice with sterile distilled water to remove any remaining media, then diluted 1:100 in 2mL SGal-ade media. The OD_{595} was monitored during growth at 30°C in a FLUOstar OPTIMA spectrophotometer. The generation time was measured as the minimum doubling time over a 30-minute interval.

Results

The E. coli def-fmt gene pair does not stabilise plasmids in S. cerevisiae

We noted during routine culturing that the pBEVY plasmids used in this work are lost at high frequencies in the absence of selection. The *ADE2* marker used in these plasmids provides a visual indicator of plasmid loss. An *ade2* strain (such as W303α) appears red when grown on rich media in the absence of adenine due to the accumulation of a pigmented intermediate in the adenine biosynthetic pathway (Chaudhuri et al., 1997). In contrast, complementation of *ade2* with an *ADE2*-encoding plasmid (such as pBEVY) results in formation of white colonies on rich media. Thus, W303α cells containing pBEVY produce white colonies on YPD or YPGal media, whereas plasmid-less W303α cells produce red colonies. When grown on media repressing expression from GAL promoters (YPD), W303α strains containing pBEVY-derived plasmids produce white colonies with irregular red sectors indicating stochastic plasmid loss (Figure 1a). When grown on media inducing expression from GAL promoters (YPGal), W303α strains containing pBEVY, pBEVY::*def* or pBEVY::*def,fmt* plasmids produce sectored colonies as frequently as when cultured on YPD (Figure 1b). The degree of colony sectoring of W303α + pBEVY::def,fmt was similar to colonies grown under repressing conditions (YPD). This suggests that the *def-fmt* gene pair does not increase the stability of plasmids in *S. cerevisiae* as it does in our DF200 *E. coli* strain (Chapter 4, Figure 3). It is worth noting that plasmids encoding *fmt* without *def* (pBEVY::*fmt*) produce sectored colonies on YPD, but almost entirely red colonies on YPGal (Figure 1b). This suggests that, without selection, pBEVY::*fmt* is lost from cultures at high frequency.

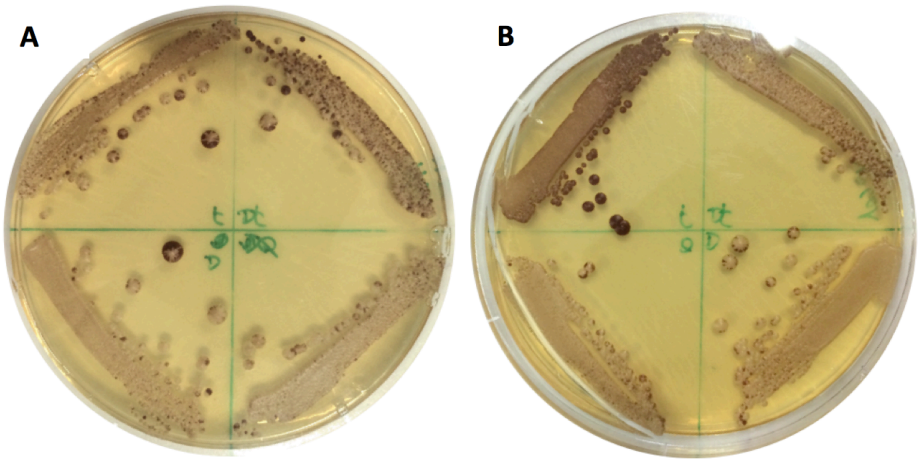


Figure 1 | The *def-fmt* gene pair does not show increased stability in *S. cerevisiae*. **A)** When grown on rich media (YPD) in the absence of adenine selection, pBEVY (and derived plasmids: pBEVY – bottom right quadrant, pBEVY::*def* – bottom left quadrant, pBEVY::*fmt* – top left quadrant, pBEVY::*def,fmt* – top right quadrant) are lost stochastically, resulting in plasmid-free cells with a red pigmentation. **B)** When grown on media to induce *def* and/or *fmt* gene expression, plasmid pBEVY::*fmt* (top left quadrant) confers a growth disadvantage, therefore pigmented, plasmid-free cells take over the colony. Plasmids pBEVY::*def,fmt* (top right quadrant), pBEVY (bottom left quadrant) and pBEVY::*def* (bottom right quadrant) are lost at similar frequencies.

Growth of S. cerevisiae is slowed by fmt expression without def

To examine the effect that cytoplasmic expression of *E. coli* formylase has on the growth of W303α, we grew cultures of W303α containing pBEVY, pBEVY::*def*, pBEVY::*fmt* or pBEVY::*def,fmt* plasmids on a solid medium under inducing conditions (SGal-ade media). To ensure cells in each of the different cultures were in the same growth phase, a sample from each culture was used to inoculate a non-inducing liquid medium (SD-ade). After these cultures reached saturation, they were washed to remove any remaining glucose (a repressor of expression from GAL promoters), and then streaked to either SD-ade or SGal-ade solid media. After 72 hours culture at 30°C, there was no visible difference in the growth of strains grown on medium (SD-ade) where the GAL1 and GAL10 promoters are repressed (Figure 2a). In contrast, when strains were grown on medium inducing the promoters (SGal-ade), W303α+pBEVY::*fmt* grew much slower than the other three strains (Figure 2b), consistent with previously observations (Ramesh et al., 2002).

To quantify the difference in growth rates observed on solid media, we monitored the growth of W303α strains in SGal-ade liquid medium (Figure 3a). The minimum doubling time of W303α expressing the *E. coli def* gene (pBEVY::*def)* was indistinguishable from strains carrying an empty vector (pBEVY) (Figure 3b). In contrast, strains expressing the *fmt* gene grew 47% slower than control strains that did not express the formylase enzyme (Figure 3b). This result confirms the slow growth phenotype observed on solid medium (Figure 2b). Deformylase expression appeared to relieve the slow growth phenotype of an *fmt*-expressing strain, resulting in a near wild-type growth rate.

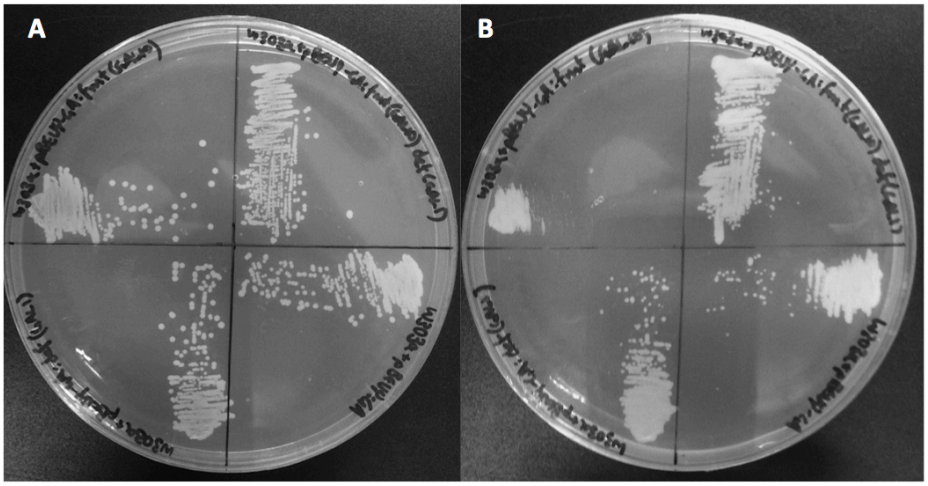


Figure 2 | *fmt* expression slows the growth of W303α. Cultures of W303α containing pBEVY, pBEVY::*def,* pBEVY::*fmt* or pBEVY::*def,fmt* were prepared in non-inducing media then streaked to either non-inducing media (SD-ade) or inducing media (SGal-ade) and grown at 30°C for 72 hours. **A)** When grown on SD-ade media, colony size appears similar for all strains. pBEVY (bottom right quadrant), pBEVY::*def* (bottom left), pBEVY::*fmt* (lop left) or pBEVY::*def,fmt* (top right). **B)** When grown on inducing media (SGal-ade), the W303α strain containing pBEVY::*fmt* (top left quadrant) grows much slower than the other strains tested.

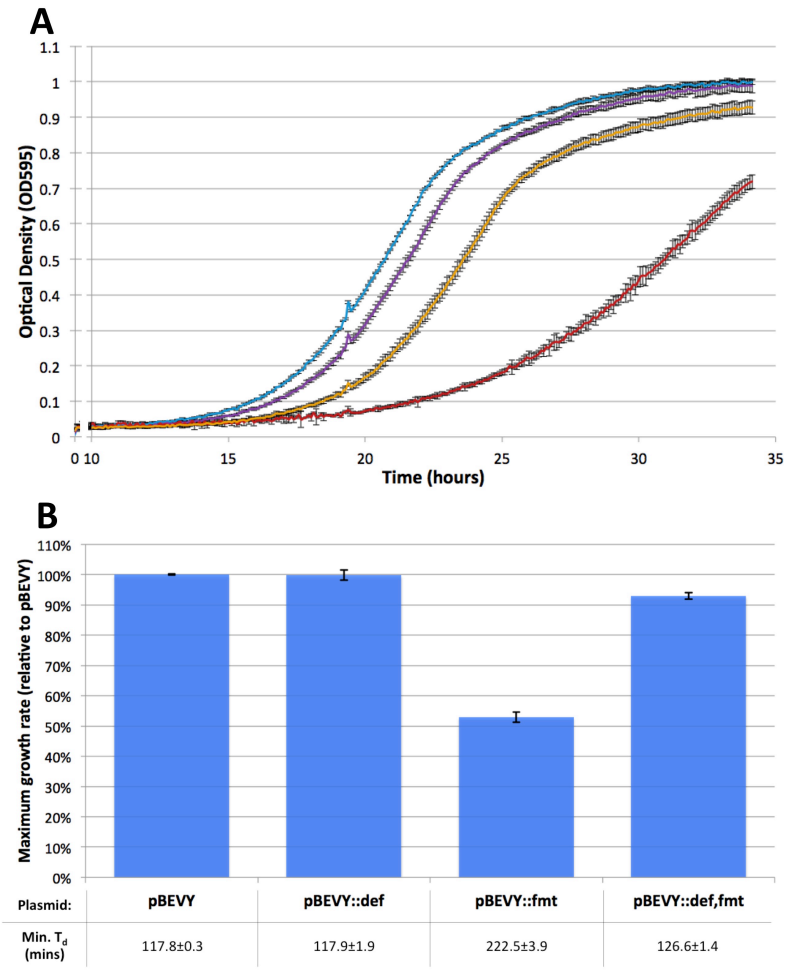


Figure 3 | *def* expression relieves the slow growth phenotype of *S. cerevisiae* expressing *fmt*. W303α strains carrying pBEVY, pBEVY::*def*, pBEVY::*fmt*, and pBEVY::*def*,*fmt*, were grown to stationary phase in SD-ade media. Cultures were washed and diluted 1:100 in SGal-ade media. OD_{595} was monitored for 34 hours. **A)** Growth curves for W303α + pBEVY (blue), pBEVY::*def* (purple), pBEVY::*fmt* (orange), and pBEVY::*def*,*fmt* (red) are shown. No growth was observed during the first 10 hours, thus this data has been removed. Points are shown as $\bar{X}$±SD, n=3. **B)** The relative growth rates were measured by calculating the minimum doubling time over a 30-minute period and comparing this to the doubling time of a strain containing a control plasmid (pBEVY). The calculated minimum doubling time in minutes (T_d) for each strain is shown below the graph. Growth rates are shown as $\bar{X}$±SE, n=3.

Actinonin does not inhibit E. coli deformylase in S. cerevisiae

We examined whether the addition of actinonin affected the growth of *S. cerevisiae* expressing the *E. coli* deformylase enzyme. This is because the *E. coli* deformylase is sensitive to inhibition by actinonin; if actinonin inhibited the enzyme when expressed in yeast, then we predicted that it would impose a cost on strains expressing both *fmt* and *def*. Actinonin was added to SGal-ade media at a concentration of 200µg/mL (10x higher than the MIC for *E. coli*) and the growth of W303α containing pBEVY plasmids measured. At this concentration, strain W303α showed no sensitivity to actinonin. As expected, yeast strains carrying an empty vector or expressing the formylase or deformylase alone showed no change in growth rate in the presence of actinonin. Surprisingly, actinonin did not cause the growth rate of strains expressing both formylase and deformylase to fall to that of a formylase-only phenotype. Given the sensitivity of *E. coli* deformylase to actinonin at below 20µg/mL (see Chapter 2, Figure 4a), we expected the growth rate of W303α+pBEVY::*def,fmt* to decrease to at least the same rate as W303α+pBEVY::*fmt* in the presence of the deformylase inhibitor. However, the growth rate of W303α+pBEVY::*def,fmt* decreased only slightly (Figure 4), maintaining a growth rate faster than formylase expressing W303α. This suggests that the addition of actinonin to yeast growth medium is not sufficient to inhibit cytoplasmic peptide deformylase activity.

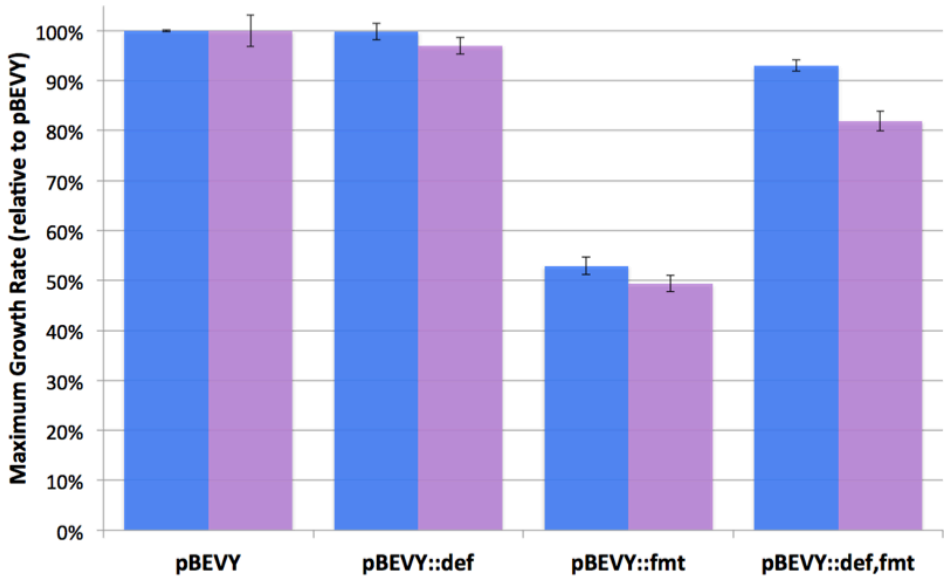


Figure 4 | *S. cerevisiae* is resistant to actinonin. W303α strains carrying pBEVY, pBEVY::*def*, pBEVY::*fmt* or pBEVY::*def-fmt* were grown in SGal-ade supplemented with 200µg/mL Actinonin. The minimum doubling time was calculated over a 30-minute interval. Growth rates are shown relative to the control strain (W303α+pBEVY) in the presence (purple bars) or absence (blue bars) of actinonin. No change in growth rate was observed for strains containing pBEVY, pBEVY::*def*, or pBEVY::*fmt*, however a small decrease in growth rate was observed for strains expressing both formylase and deformylase (pBEVY::*def-fmt*). Bars represent $\bar{X}\pm$SE, n=3.

E. coli def and fmt genes are expressed in S. cerevisiae

As the *def-fmt* gene pair results in stabilisation of a plasmid in our *E. coli* DF200 strain, but not in *S. cerevisiae* we examined whether the *def* and *fmt* genes are indeed expressed in yeast. To confirm the activity of the GAL1 and GAL10 promoters in W303α, we used RT-PCR to identify the presence of *def* and/or *fmt* transcripts in galactose-supplemented cultures. W303α strains containing plasmids which encode the *E. coli fmt* gene i.e. pBEVY::*fmt* and pBEVY::*def,fmt* show detectable transcripts for *fmt* (Figure 5a). A similar result was observed for strains containing plasmids which encode the *E. coli def* gene i.e. pBEVY::*def* and pBEVY::*def,fmt*, where *def* transcripts were detectable (Figure 5b). No *fmt* transcripts were detectable in strains without *fmt*-encoding plasmids and no *def* transcripts were detectable in strains without *def*-encoding plasmids. This confirms that the RT-PCR was specific to the *E. coli def* and *fmt* sequences. None of our reactions show any amplification product in controls where the reverse transcriptase was absent (Figure 5, lanes RT: –), indicating that any contaminating DNA was below the threshold concentration for detectable amplification. Neither *def* nor *fmt* transcripts were amplified from extracts of strain W303α+pBEVY, as would be expected for a strain without the *E. coli def* or *fmt* gene. However, this lack of amplification could also result from a failure to isolate RNA. To confirm the presence of RNA in all of our samples, especially W303α+pBEVY, we carried out a further RT-PCR to detect transcripts from the *ADE2* gene – the marker gene of the pBEVY-GA plasmids. *ADE2* transcripts were detectable in all plasmid-bearing strains (Figure 5c), confirming the stability of RNA in the preparations and suggesting that the absence of *def* and *fmt* RT-PCR products was due to the absence of *def* and *fmt* transcripts.

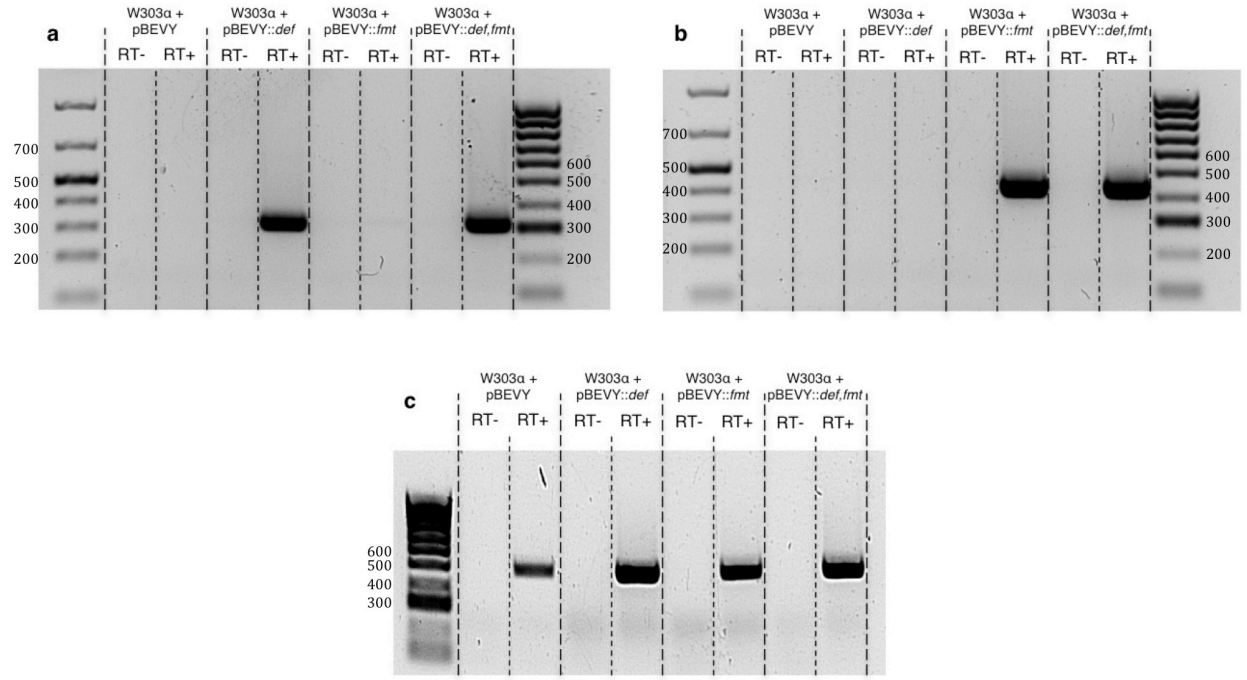


Figure 5 | RT-PCR confirms *E. coli def* and *fmt* genes are expressed in *S. cerevisiae*. Total RNA was extracted from early stationary phase cultures of *S. cerevisiae* strain W303α containing pBEVY-derived plasmids and used as a template for RT-PCR. RT- lanes were processed without a reverse transcriptase reaction in order to detect the presence of plasmid DNA. RT+ lanes were processed as per the manufacturer's guidelines. **a)** RT-PCR with primers corresponding to *E. coli def* gene (expected product size 319bp). Lane 1, GeneRuler 1kb Plus (Thermo Scientific) with size standards indicated; lanes 2-9, RT-PCR products for W303α+pBEVY, pBEVY::*def,* pBEVY::*fmt* and pBEVY::*def,fmt* with RT- controls indicated; lane 10, Hyperladder 4 (Bioline) with size standards indicated. Positive amplification observed for W303α+pBEVY::*def* and W303α+pBEVY::*def,fmt.* **b)** RT-PCR with primers corresponding to *E. coli fmt* gene (expected product size 440bp). Lanes as described in a. Postive amplification observed for W303α+pBEVY::*fmt* and W303α+pBEVY::*def,fmt.* **c)** RT-PCR with primers corresponding to *ADE2* (expected product size 497bp). Lane 1, Hyperladder 4 (Bioline) with size standards indicated; lanes 2-9, RT-PCR products for W303α+pBEVY, pBEVY::*def,* pBEVY::*fmt* and pBEVY::*def,fmt* with RT- controls indicated. Positive amplification observed for all samples.

Discussion

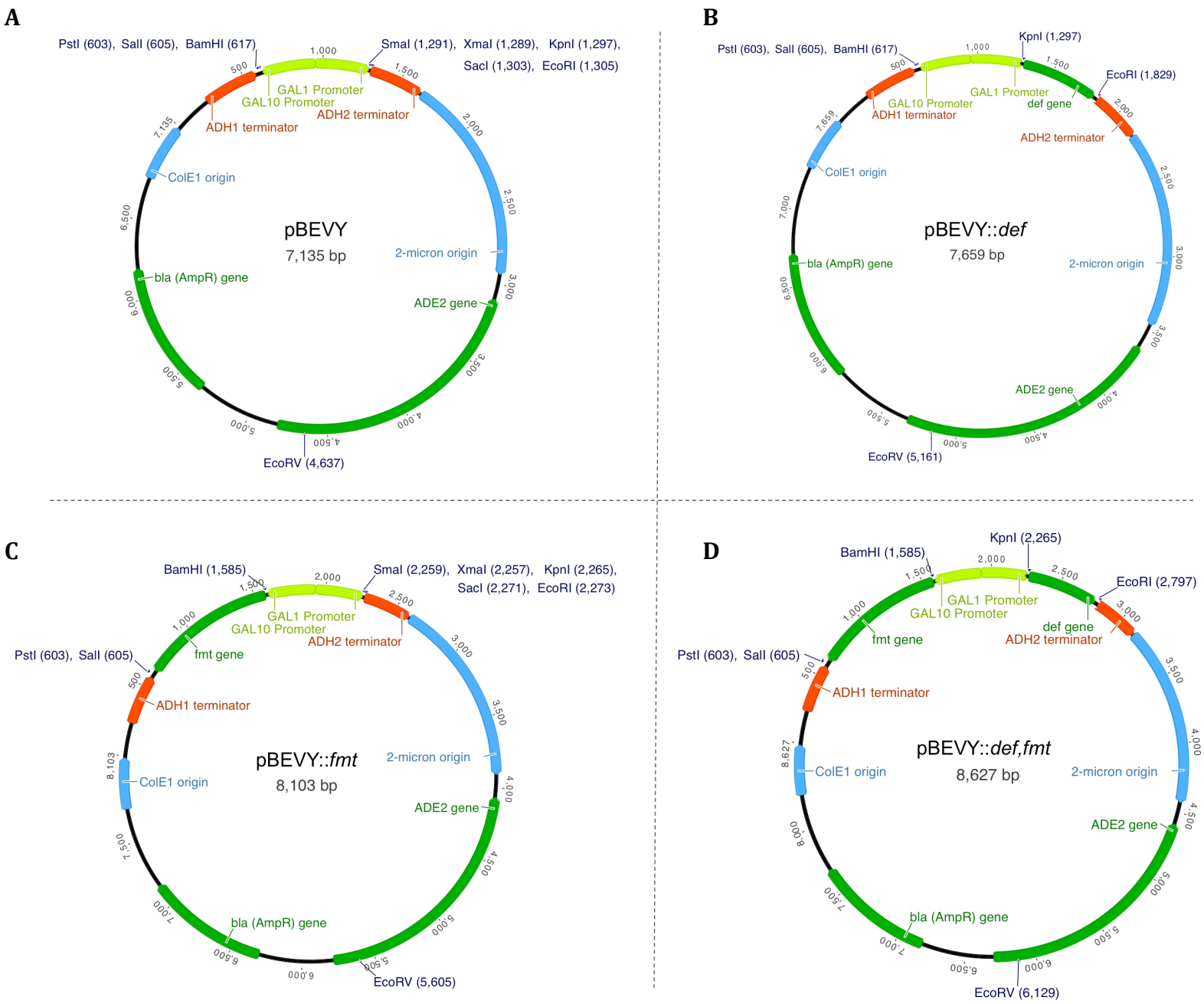
We have shown previously that Met-tRNA$_i^{fMet}$ formylation is lethal in the absence of deformylation in *E. coli* (Chapter 2, Figure 4; Chapter 3). In contrast, our results clearly indicate that formylation of the Met-tRNA$_i^{Met}$ is not lethal in *S. cerevisiae,* even in the absence of deformylation. This difference in the toxicity of formylation could result from several causes:

1. *E. coli* formylase may not be active in *S. cerevisiae*. We were unable to detect *S. cerevisiae* tRNA$_i$ by northern blotting, thus unable to confirm the presence of cytoplasmic fMet-tRNA$_i^{Met}$. However, the recovery of growth rate observed when *E. coli def* was expressed in parallel with *fmt,* suggests formylated Met-tRNA$_i^{Met}$ was not only present, but used in translation initiation in W303α+pBEVY::*def,fmt*. Peptide deformylase cleaves the formyl-moiety from the *N*-terminus of nascent polypeptides and is unable to hydrolyse formylated initiator tRNA, hence the recovery of growth rate we observed is consistent with the deformylation of peptides. The deformylation of peptides is only a requirement if fMet is used as the initiating amino acid in translation. Additionally, this use of fMet-tRNA$_i^{Met}$ in *S. cerevisiae* is consistent with previous observations (Ramesh et al., 2002).
2. *S. cerevisiae* may be able to tolerate a formylated proteome, albeit at a reduced growth rate. No data is available regarding formylation directly, however it is clear that methionine aminopeptidase (MAP) activity is essential in *S. cerevisiae* (Li and Chang, 1995). As MAP is unable to hydrolyse *N*-blocked methionine from peptides in bacteria (Solbiati et al., 1999), formylated proteins may not be processed by MAP in *S. cerevisiae*. Consequently, the essential MAP activity would be blocked, resulting in cell death. Thus, the viability of *S. cerevisiae* expressing *fmt* is unlikely due to tolerance of a formylated proteome.

3. *S. cerevisiae* may possess peptide deformylase activity. A weak deformylase activity may facilitate growth in the presence of formylase activity. Indeed, previous studies have noted peptide deformylase activity in eukaryotic cytoplasmic extracts (Yoshida and Lin, 1972), however it is difficult to determine whether this activity is true cytoplasmic deformylase or a result of mitochondrial deformylase. If such activity exists in *S. cerevisiae,* it must be less efficient than the *E. coli def* expressed from the *GAL1* promoter, as *def* expression increases the growth rate of *fmt* expressing cells.
4. Aminopeptidase enzymes which cleave *N*-formylmethionine may be present in *S. cerevisiae.* Enzymes which hydrolyse *N*-blocked amino acids from proteins have been identified in many eukaryotes (Farries et al., 1991; Gade and Brown, 1978; Jones et al., 1987; Nguyen and Pei, 2005; Radhakrishna and Wold, 1989). These enzymes are termed acylaminoacyl-peptidases (AAP) (EC 3.4.19.1) due to their propensity to release acyl-amino acids acids from acylated peptides. However, AAPs are not specific to acylated peptides, hydrolysing formylmethionine from *N*-formylated peptides, albeit at a lower efficiency (Radhakrishna and Wold, 1989). If an AAP enzyme (or indeed any similarly functioning enzyme) were present in *S. cerevisiae*, it may have sufficient activity to allow the survival of cells expressing *fmt* and subsequently producing formylated proteins.

Regardless of the mechanism involved, it is clear that *S. cerevisiae* is able to tolerate expression of *E. coli fmt* in the absence of *def*-expression. We showed previously that formylation activity is toxic in *E. coli,* with deformylation being antidotal. This toxin-antitoxin activity is the basis of PSK elicited by *def-fmt*-encoding plasmids in our *E. coli* DF200 strain (see Chapter 4). The resistance of *S. cerevisiae* to the toxicity of formylation would render the lethal activity of PSK from the *def-fmt* gene pair ineffective. If formylase and deformylase activity were lost from yeast cytoplasmic translation, the lability of deformylase activity relative to formylase activity may result in the production of formylated polypeptides.

However, this is not lethal in *S. cerevisiae*. Consequently, the addictive, plasmid stabilising phenotype displayed by the *def-fmt* operon in our DF200 strain is absent in yeast. Without PSK activity, *def-fmt* gene pairs transferred to the nuclear genome through HGT or EGT may show fleeting adoption in cytoplasmic translation, however these activities could easily be lost. As a result, cytoplasmic translation could be restored to a non-formylating state without causing cell death. The mitochondrial and plastid ancestors, having descended from bacterial endosymbionts, rely on formylation in translation. Consequently, the loss of formylase and deformylase activity would elicit a PSK effect, as observed in bacteria (Chapter 4). Thus, the *def-fmt* gene pair must be either maintained in mitochondrial/plastid genomes, or, following transfer to the nucleus, formylase and deformylase enzymes must be targeted to the required organelles.



Supplementary Figure 1: Plasmids used for *def* and *fmt* expression in *S. cerevisiae*.

A) pBEVY plasmid (Genbank Accession: AF069718) obtained from C. Miller.

B) *def* gene from *E. coli* strain REL606 amplified with a synthesized GCCACCATG Kozak sequence and inserted between *Kpn*I and *EcoR*I sites of GAL1 promoter MCS to produce pBEVY::*def*.

C) *fmt* gene from *E. coli* strain REL606 amplified with a synthesized GCCACCATG Kozak sequence and inserted between *BamH*I and *Sal*I sites of GAL10 promoter MCS. Start codon of *fmt* is modified from the native GTG to ATG to facilitate expression in yeast.

D) Both the *def* and *fmt* genes inserted in to pBEVY to produce the plasmid pBEVY::*def,fmt*.

References

Andersson, J.O. (2005). Lateral gene transfer in eukaryotes. Cell. Mol. Life Sci. *62*, 1182–1197.

Becker, D.M., and Guarente, L. (1991). High-efficiency transformation of yeast by electroporation. C. Guthrie, and G. Fink, eds. (Elsevier), pp. 182–187.

Carrie, C., and Small, I. (2013). A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research *1833*, 253–259.

Chaudhuri, B., Ingavale, S., and Bachhawat, A.K. (1997). apd1+, a Gene Required for Red Pigment Formation in ade6 Mutants of Schizosaccharomyces pombe, Encodes an Enzyme Required for Glutathione Biosynthesis: A Role for Glutathione and a Glutathione-Conjugate Pump. Genetics *145*, 75–83.

Farries, T.C., Harris, A., Auffret, A.D., and Aitken, A. (1991). Removal of N-acetyl groups from blocked peptides with acylpeptide hydrolase. European Journal of Biochemistry *196*, 679–685.

Gade, W., and Brown, J.L. (1978). Purification and partial characterization of alpha-N-acylpeptide hydrolase from bovine liver. J. Biol. Chem. *253*, 5012–5018.

Garofalo, C., Trinko, R., Kramer, G., Appling, D.R., and Hardesty, B. (2003). Purification and characterization of yeast mitochondrial initiation factor 2. Arch. Biochem. Biophys. *413*, 243–252.

Giglione, C. (2000). Identification of eukaryotic peptide deformylases reveals universality of N-terminal protein processing mechanisms. EMBO J. *19*, 5916–5929.

Giglione, C., and Meinnel, T. (2001). Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. Trends Plant Sci. *6*, 566–572.

Heinemann, J.A., and Kurenbach, B. (2009). Horizontal transfer of genes between microorganisms. In Encyclopedia of Microbiology, (Elsevier Inc.).

Heinemann, J.A., and Sprague, G.F. (1989). Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature *340*, 205–209.

Huang, C.Y., Ayliffe, M.A., and Timmis, J.N. (2003). Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature *422*, 72–76.

Jones, W.M., Manning, L.R., and Manning, J.M. (1987). Properties of the Hydrolase that Catalyzes Removal of the Blocked NH2-Terminal Amino Acid

Residues from Polypeptides. In Proteins, (Boston, MA: Springer US), pp. 675–681.

Li, X., and Chang, Y.H. (1995). Amino-terminal protein processing in Saccharomyces cerevisiae is an essential function that requires two distinct methionine aminopeptidases. PNAS *92*, 12357–12361.

Liao, H.X., and Spremulli, L.L. (1991). Initiation of protein synthesis in animal mitochondria. Purification and characterization of translational initiation factor 2. J. Biol. Chem. *266*, 20714–20719.

Marcker, K., and Sanger, F. (1964). N-Formyl-methionyl-S-RNA. J. Mol. Biol. *8*, 835–IN838.

Naveau, M., Lazennec-Schurdevin, C., Panvert, M., Mechulam, Y., and Schmitt, E. (2010). tRNA binding properties of eukaryotic translation initiation factor 2 from Encephalitozoon cuniculi. Biochemistry *49*, 8680–8688.

Nguyen, K.T., and Pei, D. (2005). Purification and characterization of enzymes involved in the degradation of chemotactic N-formyl peptides. Biochemistry *44*, 8514–8522.

Radhakrishna, G., and Wold, F. (1989). Purification and characterization of an N-acylaminoacyl-peptide hydrolase from rabbit muscle. J. Biol. Chem. *264*, 11076–11081.

Ramesh, V., Köhrer, C., and RajBhandary, U.L. (2002). Expression of *Escherichia coli* methionyl-tRNA formyltransferase in *Saccharomyces cerevisiae* leads to formylation of the cytoplasmic initiator tRNA and possibly to initiation of protein synthesis with formylmethionine. Mol. Cell. Biol. *22*, 5434–5442.

Ream, W. (1989). Agrobacterium tumefaciens and interkingdom genetic exchange. Annu. Rev. Phytopathol. *27,* 583-618.

Schmitt, M.E., Brown, T.A., and Trumpower, B.L. (1990). A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic Acids Res. *18*, 3091–3092.

Smith, A.E., and Marcker, K.A. (1968). N-formylmethionyl transfer RNA in mitochondria from yeast and rat liver. J. Mol. Biol. *38*, 241–243.

Smith, A.E., and Marcker, K.A. (1970). Cytoplasmic methionine transfer RNAs from eukaryotes. Nature *226*, 607–610.

Solbiati, J., Chapman-Smith, A., Miller, J.L., Miller, C.G., and Cronan, J.E., Jr (1999). Processing of the N termini of nascent polypeptide chains requires

deformylation prior to methionine removal. J. Mol. Biol. *290*, 607–614.

Stegemann, S., Hartmann, S., Ruf, S., and Bock, R. (2003). High-frequency gene transfer from the chloroplast genome to the nucleus. PNAS *100*, 8828–8833.

Takeuchi, N., Kawakami, M., Omori, A., Ueda, T., Spremulli, L.L., and Watanabe, K. (1998). Mammalian Mitochondrial Methionyl-tRNA Transformylase from Bovine Liver. J. Biol. Chem. *273*, 15085–15090.

Yoshida, A., and Lin, M. (1972). NH2-Terminal Formylmethionine- and NH2-Terminal Methionine-Cleaving Enzymes in Rabbits. J. Biol. Chem. *247,* 952-957.

Chapter 6

Evidence for a role of formyl-methionine use in Archaea

All experimental work outlined in this chapter was performed by Ryan Catchpole under the mentorship of Prof. Jack Heinemann and Assoc. Prof. Anthony Poole. Bioinformatic analyses were performed by Ryan Catchpole with the assistance of Dr. Stinus Lindgreen.

Introduction

Archaea initiate translation using a non-formylated methionyl-initiator tRNA (Met-$tRNA_i^{Met}$) (Ramesh and RajBhandary, 2001; White and Bayley, 1972a). This is in contrast to almost all known bacteria, where translation initiation proceeds with a formylated initiator tRNA (fMet-$tRNA_i^{fMet}$) (Kozak, 1983). The apparent absence of methionyl-tRNA formyltransferase and peptide deformylase genes from the archaeal domain is surprising, especially in light of extensive horizontal gene transfer between bacteria and archaea (Aravind et al., 1998; Brochier-Armanet and Forterre, 2007; Nelson et al., 1999). Additionally, the toxin-antitoxin phenotype shown by the *def-fmt* gene pair in our DF200 lines (see Chapter 3) indicates that these genes may have an increased propensity to spread through horizontal gene transfer.

At the time of publication, 292 archaeal genomes were available in the NCBI database, with 67% (197) of these from a single phylum (Euryarchaeota), and 33% (97) from a single family (Halobacteriaceae). In contrast, 4,357 bacterial genomes were available across 26 phyla. The highly skewed nature of the archaeal genome

dataset limits the diversity of data available for analysis. Next-generation sequencing of metagenomic isolates is now providing a large amount of sequence data for unculturable archaea, however with limited data available on archaeal genes and few biochemical studies, annotation of these metagenomic sequences is difficult.

We asked whether there is evidence for formylase or deformylase genes in archaeal genomes. We used several bioinformatics methods to search for and characterize archaeal genes with similarity to bacterial formylase and deformylase genes. We provide evidence for the presence of a putative *def-fmt* operon in the two representative genomes of the archaeal *Methanocorpusculum* genus. We also identify a putative *def* gene in the genome of a marine *Halococcus* species, however further analysis shows this to be a sequencing artefact. The data we present provides evidence for the presence of methionyl-tRNA formyltransferase and peptide deformylase in the archaeal domain and challenges the current belief that fMet-tRNA$_i$ is limited to bacteria and bacteria-derived organelles of eukaryotes.

Methods

Sequence search

pHMMER searches were carried out using the HMMER web server (Finn et al., 2011) (http://hmmer.janelia.org/). Default parameters were used unless specified.

BLASTp searches were carried out using the BLAST web server (Altschul et al., 1990) (http://blast.ncbi.nlm.nih.gov). Default parameters were used unless specified.

Phylogenetic trees

Sequences used to generate phylogenetic trees were obtained from a BLASTp search against a local copy of the non-redundant database (March 2012). Full-length amino-acid sequences were used as query sequences and the top 150 hits were selected based on the lowest E-value (maximum limit set at 0.01). Each set of hits was aligned with MAFFT (Katoh et al., 2002), MUSCLE (Edgar, 2004) and ClustalW (Larkin et al., 2007). To determine which program produced the highest quality alignments, each was analysed with the T-COFFEE (Notredame et al., 2000) web server (http://tcoffee.crg.cat). MUSCLE alignments consistently gave quality scores higher than or equal to the MAFFT and ClustalW alignments, and thus were used for further analysis.

To determine the most appropriate amino acid replacement rate matrix for tree generation, alignments were analyzed with ProtTest (Abascal et al., 2005). Alignments based on Mlab_191 and Mlab_192 were predicted to have a 'mtmam' substitution model. This model is based on mitochondrial sequences (Yang et al., 1998), and may thus be inappropriate for our bacterial and archaeal sequences. As a result, we built and compared trees using multiple substitution models. Trees were generated in PhyML (Guindon et al., 2010) using the models JTT, LG, WAG and mtmam; and in MrBayes (Huelsenbeck and Ronquist, 2001) using Jones, Dayhoff, WAG and mtmam.

In order to remove poorly-aligned regions from our dataset, MUSCLE alignments were also processed using G-Blocks (Castresana, 2000), then used to generate trees as described above. These trees showed no noTable differences in topology to those generated using raw alignment data, and thus were discarded.

Structural prediction

Homology models and structural predictions were generated using the Phyre2 (Kelley and Sternberg, 2009) web server (http://www.sbg.bio.ic.ac.uk/phyre2/).

Halococcus hamelinensis

H. hamelinensis 100A6 was obtained from Dr. Brendan Burns (University of New South Wales, Australia) and also from the Japan Collection of Microorganisms (JCM 12892). Strains were grown on solid DSM-97 agar media (DasSarma et al. 1995) at 37°C for 7 days. Colonies were scraped from agar plates and vortexed with glass beads (Sigma G8772) to lyse cells. Genomic DNA isolated using the Wizard Genomic DNA Purification Kit (Promega).

Primers used for PCR are listed below:

H. hamel. def1 - 5′ – CGATGACATGTTGGAAACCATGTACG – 3′

H. hamel. def2 - 5′ – GTGATCATCTGCCGCTTCAGC – 3′

H. hamel. def3 - 5′ – CCAAGGTGGTCGAGGTCGAATGG – 3′

H. hamel. def4 - 5′ – GGCGGGTCACCTCGGCGTATTGG – 3′

H. hamel. def5 - 5′ – GTGCTCGATTGCGTCAAGGAAGAGG – 3′

Genome mapping was carried out using Geneious v6.1.6 (Biomatters).

Results

Methanocorpusculum labreanum encodes putative formylase and deformylase genes

Given that *def* and *fmt* genes are found in both the bacterial and eukaryotic domains (in mitochondria and plastids) (Kozak, 1983), we performed a bioinformatic search to look for evidence of *def* and *fmt* genes in archaea. The *def* sequence from *E. coli* strain MG1655 (AAC76312.1) was used as the query sequence for a pHMMER search against the UniProt Knowledgebase (UniProtKB). Two significant hits (significance limited to E-values below 0.01) were found in archaea, both from *Methanocorpusculum labreanum.* These hits were to two proteins annotated as 'peptide deformylase' with E-values of $8.5x10^{-19}$ and $3.0x10^{-12}$ (accession numbers ABN06367.1 and ABN06388.1, respectively). Performing the same search with the *E. coli fmt* sequence (P23882.4) gave 71 significant hits – the number of hits was likely inflated by sequence similarities of tetrahydrofolate (a cofactor for formylase) binding domains. The hit of greatest significance (E-value of $4.1x10^{-67}$) was another protein from *M. labreanum* annotated as 'methionyl-tRNA formyltransferase' (ABN06368.1).

Intriguingly, two of the hits are encoded by adjacent genes in the *M. labreanum* genome (NC_008942.1), separated by a single nucleotide. These genes, Mlab_191 and Mlab_192, encode the annotated peptide deformylase and methionyl-tRNA formyltransferase, respectively. This arrangement fits with our putative *def-fmt* operon arrangement (Chapter 4, Figure 1), with the deformylase gene encoded immediately upstream of the adjacent formylase gene.

The annotations given to Mlab_191 and Mlab_192 following genome sequencing were based on automatic annotation software (Anderson et al., 2009). To examine

the reliability of these annotations, we carried out further analysis of these protein sequences.

M. labreanum sequences show predicted structural similarity to bacterial formylase and deformylase enzymes

We observed that the presence of co-factor binding domains e.g. tetrahydrofolate can result in hits with significant E-values, despite proteins having different annotated activities. To test whether the *M. labreanum* peptide deformylase and methionyl-tRNA formyltransferase hits were likely to show structural similarity to known deformylase and formylase enzymes, we used Phyre2 to generate structural homology models of these ORFs. In all three cases, each ORF sequence was modelled with 100% confidence to known formylase and deformylase enzymes (Table 1, Table 2 & Table 3). This indicates that, not only do these predicted proteins show sequence similarity to known formylase and deformylase genes, they also show structural similarity to functional enzymes (Figure 1). This suggests that these genes could function in the same way as bacterial formylase and deformylase enzymes.

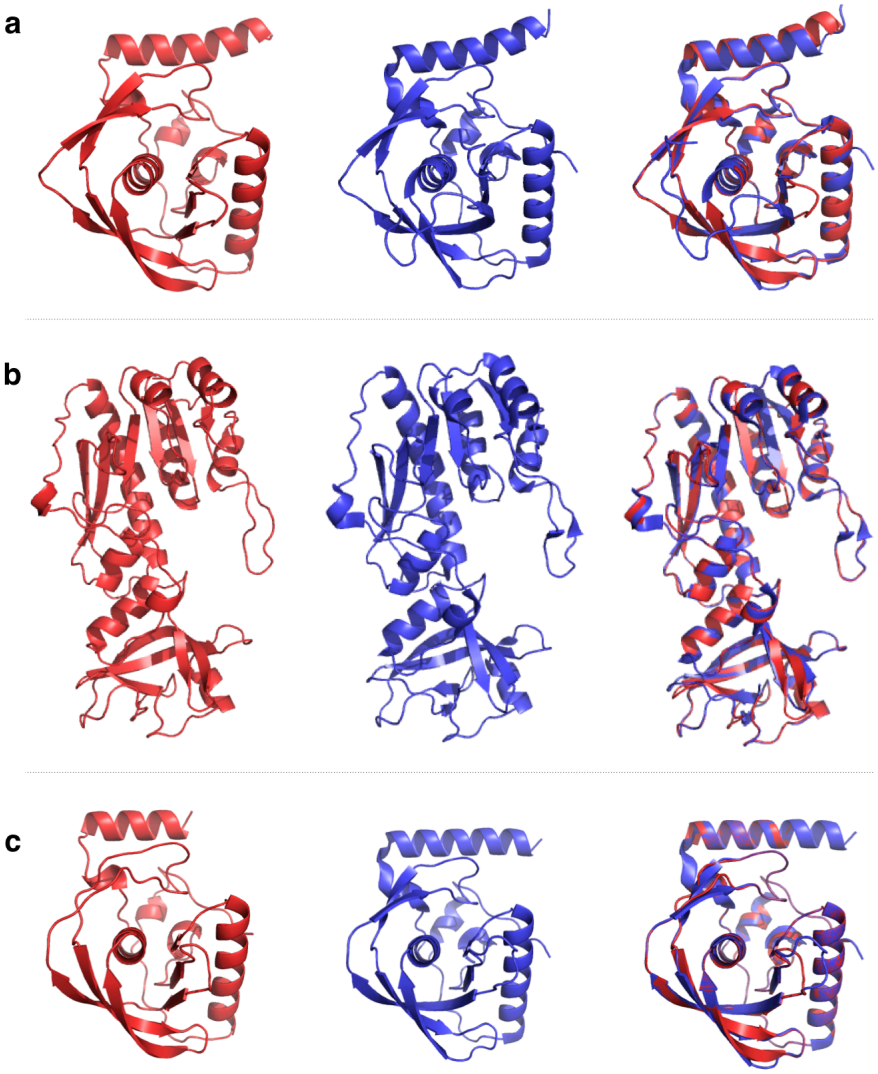


Figure 1 | Sequences from *M. labreanum* show predicted structural similarity to known formylase and deformylase enzymes. Phyre2 was used to generate structural homology models for three ORFs from *M. labreanum.* **a)** Predicted structure of Mlab_191 (red, left) shows similarity to peptide deformylase from *Ehrlichia chaffeensis* (3oca) (blue, center). An alignment of the two structures (right) shows they are near identical. **b)** Predicted structure of Mlab_192 (red, left) shows similarity to methionyl-tRNA formyltransferase from *Bacillus anthracis* (3rfo) (blue, center). An alignment of the two structures (right) shows they are near identical. **c)** Predicted structure of Mlab_212 (red, left) shows similarity to peptide deformylase from *Escherichia coli* (2w3t) (blue, center). An alignment of the two structures (right) shows they are near identical. Images generated in PyMOL (Schrödinger LLC).

M. labreanum tRNA$_i$ shows features required for formylation

A functional formylase enzyme is ineffective without an appropriate $tRNA_i$ substrate. Specific base pairing features of the initiator tRNA in bacteria facilitate recognition by the formylase enzyme and promote formylation of the bound methionine (Lee et al., 1991). In particular, formylation is dependent on a weak base pair (e.g. A:U) or a mismatch at the top of the acceptor stem, as well as bases equivalent to A73, G2:C71, C3:G70 and G4:C69 of the *E. coli* $tRNA_i^{fmet}$. To examine whether these features were present in *M. labreanum* $tRNA_i^{Met}$, we used tRNAScan-SE (Lowe and Eddy, 1997) to search the *M. labreanum* genome (NC_008942.1) for tRNA sequences. Three Met-tRNA sequences were identified, however only one of these sequences showed all of the features conserved in archaeal initiator tRNAs (Ushida et al., 1996). This $tRNA_i$ sequence also showed all of the bases required for recognition by the formylase enzyme (Figure 2).

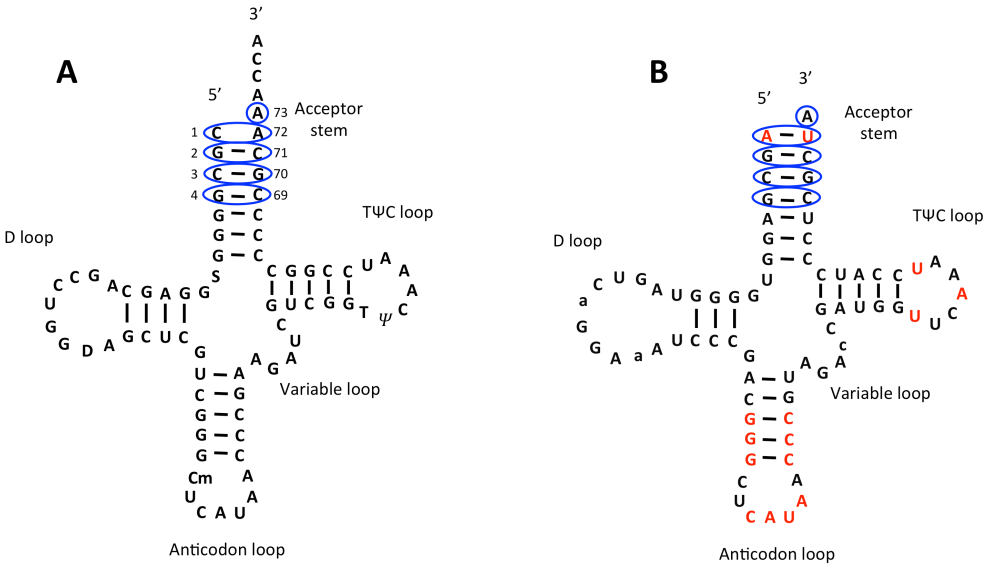


Figure 2 | *M. labreanum* initiator tRNA shows features important for formylation in *E. coli*. A) Sequence of the *E. coli* initiator tRNA with bases/pairings important for formylation circled in blue. Base pairs G2:C71, C3:G70 and G4:C69, as well as the unpaired A73 and a non-G:C base pair at positions 1:72 promote formylation by methionyl-tRNA formyltransferase. B) The putative initiator tRNA identified in *M. labreanum*. Bases labeled in red are conserved in archaeal initiator tRNAs and were used to identify the initiator tRNA over elongator Met-tRNAs. This tRNA possesses all of the features (circled in blue) important for formylation of Met-tRNA$_i^{fmet}$ in *E. coli*.

Archaeal formylase and deformylase-like sequences may have a bacterial origin

To determine whether the presence of formylase and deformylase genes in the *M. labreanum* genome is the result of a bacteria-to-archaea horizontal gene transfer, or whether the genes had evolved independently from other archaeal genes, we examined the similarity of these genes to other sequences. We used the *M. labreanum* ORF sequences Mlab_191 and Mlab_192 as query sequences for a BLASTp search against a local version of the non-redundant database (downloaded March 2012). The top 150 hits (based on E-value) were used to build phylogenetic trees. If Mlab_191 and Mlab_192 had evolved from archaeal genes, we would expect them to group most closely with other archaeal sequences in a phylogenetic tree. Rather, the *M. labreanum* sequences group most closely with bacterial sequences (Supplementary Figures 4&5) with the two deformylase protein sequences and the formylase protein sequence falling within a clade containing members of the order Fusobacteriales e.g. *Fusobacterium, Ilyobacter* and *Leptotrichia.* Additionally, the two deformylase proteins (encoded by Mlab_191 and Mlab_212) group most closely with each other rather than with other bacterial sequences (Supplementary Figure 4). Despite low pairwise identity (50.7% pairwise identity at the nucleotide level and 46.8% identity at the amino acid level), this may suggest that this pair of genes arose through a duplication event within *M. labreanum.* Alternatively, these genes may have arisen through independent horizontal gene transfer events involving similar but as yet unidentified donors.

The grouping of Mlab_191 and Mlab_192 within a similar clade of organisms, despite showing no homology to one another, suggests that these genes have the same evolutionary origin, transferring from bacteria to archaea in a single horizontal transfer event. To test whether this transfer extended to genes surrounding the *def-fmt* gene pair, we constructed phylogenetic trees for the genes

surrounding Mlab_191 and Mlab_192. The three upstream genes (Mlab_188, Mlab_189 and Mlab_190) (Supplementary Figures 1, 2 & 3) and the three downstream genes (Mlab_193, Mlab_194 and Mlab_195) (Supplementary Figures 6, 7 & 8) do not group in the same way as the deformylase and formylase genes. Rather, these genes group with sequences from other methanogenic archaea, suggesting they evolved in an archaeal ancestor. This finding suggests that Mlab_191 and Mlab_192 transferred to the genome of a *M. labreanum* ancestor as a gene pair, inserting between the archaeal genes Mlab_190 and Mlab_193.

Differences in GC content can be used to delineate boundaries of recent gene transfer events (Garcia-Vallvé et al., 2000). To examine whether the boundaries of the Mlab_191-Mlab_192 gene transfer can be discerned, we measured the GC content of the *M. labreanum* Mlab_187-Mlab_195 genome region using a 100bp-sliding window. There are no distinct differences in GC content between Mlab_190 and Mlab_191 or Mlab_192 and Mlab_193 that may indicate archaeal and bacterial sequence boundaries (Figure 3).

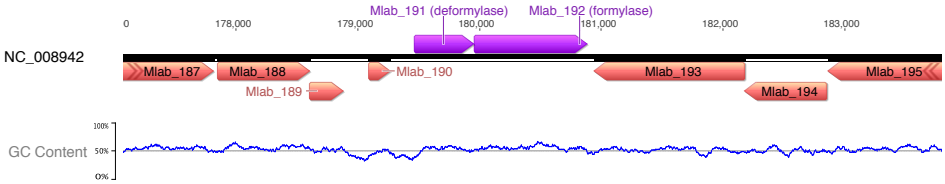


Figure 3 | GC content in *def-fmt* region of *M. labreanum* genome. The GC content (shown in blue) of a 100bp sliding window corresponding to the Mlab_187 – Mlab_195 region of the *M. labreanum* genome was plotted. Mlab_191 and 192 together have a GC content of 55.1%, only slightly higher than the 50.0% observed for the entire genome.

M. bavaricum also shows formylase and deformylase-like sequences

During the course of this work, a second *Methanocorpusculum* genome sequence was published. This genome was from the species *M. bavaricum,* isolated in Regensburg, Germany (Zellner et al., 1989) (*M. labreanum* was isolated in Los Angeles, USA (Zhao et al., 1989)). *M. bavaricum* showed the presence of a deformylase and formylase gene pair, with amino acid identities of 89.5% and 85.5% relative to *M. labreanum.* A visual comparison of the genomic region surrounding the *def-fmt* gene pair showed the presence of ORFs corresponding to Mlab_188-Mlab_195 (Figure 4). Each of these ORFs were among the top BLAST hits to *M. labreanum* (Table 5).

To determine the extent of synteny between these genomes, an alignment was performed using Mauve (Darling et al., 2010). The synteny displayed between this region of the *M. bavaricum* genome and that of *M. labreanum* extends for 27kb. Such similarity indicates that the acquisition of this *def-fmt* gene pair by these two species was likely a single horizontal transfer to a common ancestor of *M. bavaricum* and *M. labreanum*, rather than two separate bacterial-to-archaea gene transfer events.

Halococcus hamelinensis has a partial deformylase-like sequence

Further evidence of peptide deformylase in archaea was observed in the genome of *Halococcus hamelinensis* (Burns et al., 2012) (AJRK00000000.1). This whole genome shotgun sequence has been assembled into 453 contigs and is a complete representation of the 3.4Mbp genome. A 6-frame translation of one of these contigs (AJRK01000282.1) revealed one frame with significant similarity to a bacterial peptide deformylase sequence (Table 6). This translation frame provides the most significant hit of all 6 frames. Moreover, a structural prediction of this translated

sequence using Phyre2 gives 100% confidence in a homology model based on peptide deformylase from *Ehrlichia chaffeensis* (Table 4).

In many bacterial genomes, the peptide deformylase gene is encoded upstream of the methionyl-tRNA formyltransferase gene (Chapter 4, Figure 1). To examine whether the *H. hamelinensis* genome encodes a complete deformylase gene, and to determine the nature of sequences immediately downstream of *def*-like contig, we isolated genomic DNA from *H. hamelinensis.* Sequencing of the 16S gene confirmed the identity of the genomic DNA as that of *H. hamelinensis.* Primers were designed to amplify and confirm the presence of the entire *def*-like contig, and to facilitate sequencing by inverse PCR. Despite repeated attempts, no amplification products of the expected length were observed.

Following completion of this work, a further whole genome shotgun sequence for *H. hamelinensis* was submitted to GenBank. This sequence (AOMB00000000.1) was assembled into 46 contigs and is a complete representation of the 3.4Mbp genome. To examine the similarity between the sets of contigs from the AJRK and AOMB sequencing projects, we mapped the original 453 contigs to the 46 recently submitted contigs. We were able to map 358 of the AJRK contigs to the AOMB sequences with only 95 sequences remaining unmapped. The *def*-like fragment (AJRK01000282.1) could not be mapped to the AOMB contigs. To examine the origin of the 95 unmapped sequences, we used each as a query sequence for a nucleotide BLAST against the non-redundant database (August 2014). 83 of the unmapped sequences (including the *def*-like contig) had highly significant similarities (E-value $<10^{-100}$) to the genome of the marine bacterium, *Phaeobacter gallaeciensis* (Martens et al., 2006; Ruiz-Ponte et al., 1998). This suggests that the original sequencing data may have contained bacterial DNA, thus the *def*-like sequence identified in this study may not have originated from *H. hamelinensis.*

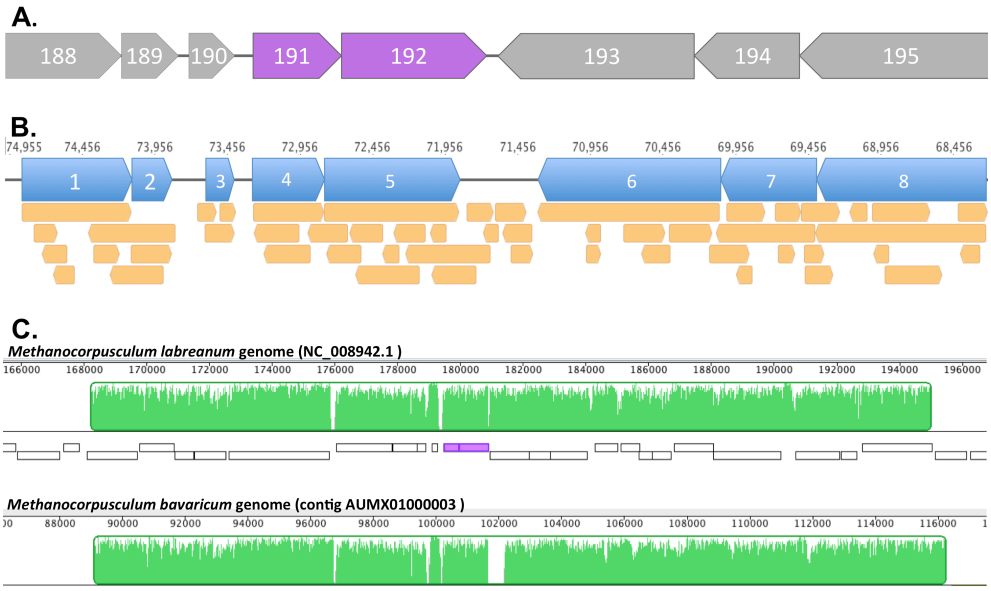


Figure 4 | The *def-fmt* region of *Methanocorpusculum labreanum* genome shows synteny with *M. bavaricum*.

A) A schematic diagram of the *M. labreanum* genome surrounding the *def-fmt* gene pair. Arrows indicate the positions of ORFs in the genome, with those in purple indicating the *def* (191) and *fmt* (192) genes. **B)** A schematic diagram of a 14kb region the *M. bavaricum* genome contig AUMX01000003 (numbers above schematic indicate base position in contig). Putative ORFs were predicted by Geneious and are indicated in orange. ORFs which show similarity in size and position to *M. labreanum* genes 188-195 are shown in blue**. C)** Synteny map produced by Progressive Mauve alignment between *M. labreanum* genome and *M. bavaricum* genome contig AUMX01000003. The region displaying synteny is labeled with green peaks where peak heights indicate the similarity of the two sequences. ORFs annotated in *M. labreanum* genome are indicated as black rectangles with the Mlab_191 and Mlab_192 labelled in purple. The region of synteny extends for 27kb.

Discussion

We have identified the presence of a *def-fmt* gene pair in the two representative genome sequences of the *Methanocorpusculum* genus. These phylogenetically similar archaea were isolated from geographically distant locations (USA and Germany) indicating that formylation may be more common in archaea than previously recognized. The sequences we have analysed likely represent a bacteria-archaea horizontal gene transfer event from a bacterium related to modern Fusobacteriales to an ancestor of *Methanocorpusculum*.

Although it is difficult to determine the functionality of genes bioinformatically, homology models of the archaeal formylase and deformylase protein structures show near identical structural prediction to known bacterial formylase and deformylase enzymes. Additionally, the *Methanocorpusculum labreanum* initiator tRNA possesses all the features important for Met-$tRNA_i^{fMet}$ formylation in bacteria, suggesting that the putative formylase may be functional in this archaeon.

The use of a formylated initiator tRNA species in archaea challenges the presumption that formylated methionine is used exclusively by bacteria is due to some idiosyncrasy in bacterial translation. Rather, archaeal initiator tRNA has the features required for formylation, and *def* and *fmt* genes are subject to horizontal gene transfer. Further work will be required to determine whether formylation of the initiator tRNA occurs *in vivo*, and whether fMet-$tRNA_i^{Met}$ is used in translation initiation.

The interdomain horizontal transfer of a *def-fmt* gene pair also challenges the predictions of the complexity hypothesis (Heinemann and Kurenbach, 2009; Jain et al., 1999). Although the *def-fmt* gene pair could be considered an independent functional unit (not known to be dependent upon gene regulatory factors (Meinnel et al., 1993)), the products and substrates of formylase and deformylase are subject to complex interactions. Recognition of the initiator tRNA by formylase requires

interactions with specific base pairs of the tRNA (Lee et al., 1991); the formylated Met-$tRNA_i$ must be recognized by initiation factor 2 and induce the appropriate conformational changes to promote ribosomal subunit assembly (Julián et al., 2011; Simonetti et al., 2008); peptide deformylase must bind adjacent to the ribosomal exit site (Bingel-Erlenmeyer et al., 2008); the difference in charge of a *N*-blocked methionine must not alter signal peptide recognition (Heijne, 1984). Consequently, horizontal transfer the *def-fmt* gene pair may alter many important protein-protein and protein-nucleic acid interactions in the cell. The extent of this interaction network is evident in the loss of fitness resulting from knocking out of formylation in bacteria (Chapter 1, Table 1), and the multiple mutations required to restore wild-type growth rate (Chapter 3, Figure 9).

The absence of known formylase and deformylase genes from archaeal genome sequences may be reinforced by the methods used to generate and annotate sequence data. Metagenomic isolates are usually processed through next-generation sequencing technologies, generating fragmented sequences. Following sequencing, fragments are assembled into longer contigs which are annotated based on similarity to known sequences. The absence of known formylase and deformylase genes from archaea combined with the ubiquity of these genes in bacteria may limit the ability of automatic annotation software to determine an archaeal origin for *def* or *fmt* containing sequence fragments. Consequently, *def-fmt* fragments from archaea may be incorrectly annotated as bacterial genes, based on the overwhelming majority of bacterial homologues. Genome sequencing projects such as the Genomic Encyclopedia of Bacteria and Archaea (GEBA) (Wu et al., 2009) are providing a wealth of data on archaeal genes. Such extensive sequencing of archaeal genomes may reveal a greater abundance of formylase and deformylase genes in the archaeal domain.

Despite the evidence that the *def-fmt* gene pair may have transferred to *Methanocorpusculum sp.*, it is clear from the genomes of other archaea that formylation is not ubiquitous in the archaeal domain. Further, biochemical studies confirm the methionyl moiety of the initiator tRNA is not formylated in *Halobacteria salinarium* (White and Bayley, 1972b). If the formylation-deformylation cycle of bacteria is as addictive as it appears (see Chapter 3), and is subject to bacteria-to-archaea horizontal gene transfer, why is formylmethionine use not prevalent in archaeal translation? A potential reason for this absence may be the presence of archaeal enzymes specific for the hydrolysis of *N*-blocked amino acids. Termed deblocking aminopeptidases, these enzymes have been identified in many hyperthermophilic archaea (Ando et al., 1999; Jia et al., 2011; Onoe et al., 2002; Story et al., 2001) and are able to hydrolyse *N*-terminal amino acids with amino-group modifications e.g. formylation and acylation. The presence of deblocking aminopeptidases would allow hydrolysis of formylmethionine from peptides, without the requirement for sequential deformylase and methionine aminopeptidase activity. This would likely abrogate the PSK activity of the *def-fmt* gene pair, allowing it to be lost from archaea. Homologues of deblocking aminopeptidase genes have been identified in *E. coli,* however, unlike archaeal relatives, these enzymes were not able to hydrolyse *N*-blocked amino acids (Zheng et al., 2005). Thus deblocking aminopeptidase activity may provide a mechanism for the scarcity of formylase and deformylase encoding genes in the archaeal domain.

Table 1 | Mlab_191 shows predicted structural similarity to peptide deformylase. The translated ORF Mlab_191 was used as a query sequence for Phyre2 structural prediction. The top four results are shown. Mlab_191 is modeled with 100% confidence to known peptide deformylase structures.

Hit Name	Hit Molecule	Confidence	% ID	PDB
Crystal structure of peptide deformylase from *Ehrlichia chaffeensis*	Peptide deformylase	100.0%	30	3oca
Crystal Structure of *P. aeruginosa* Peptide deformylase complexed with antibiotic Actinonin	Peptide deformylase	100.0%	34	1ix1
Plant peptide deformylase PDF1B crystal structure	Peptide deformylase	100.0%	35	3cpm
Chloro complex of the Ni-form of *E. coli* deformylase	Peptide deformylase	100.0%	33	2w3p

Table 2 | Mlab_192 shows predicted structural similarity to methionyl-tRNA formyltransferase. The translated ORF Mlab_192 was used as a query sequence for Phyre2 structural prediction. The top four results are shown. Mlab_192 is modeled with 100% confidence to known methionyl-tRNA formyltransferase structures.

Hit Name	Hit Molecule	Confidence	% ID	PDB
Crystal Structure of Methionyl-tRNA Formyltransferase from *Bacillus anthracis*	Methionyl-tRNA formyltransferase	100.0%	38	3rfo
Methionyl-tRNA^{fMet} formyltransferase from *Escherichia coli*	Methionyl-tRNA^{fMet} formyltransferase	100.0%	41	1fmt
Methionyl-tRNA formyltransferase from *Vibrio chloerae*	Methionyl-tRNA formyltransferase	100.0%	38	3q0i
Structure of the methionyl-tRNA formyltransferase (fmt) from *Coxiella burnetii*	Methionyl-tRNA formyltransferase	100.0%	36	3tqq

Table 3 | Mlab_212 shows predicted structural similarity to peptide deformylase. The translated ORF Mlab_212 was used as a query sequence for Phyre2 structural prediction. The top four results are shown. Mlab_212 is modeled with 100% confidence to known peptide deformylase structures.

Hit Name	Hit Molecule	Confidence	% ID	PDB
Chloro complex of the Ni-form of *E. coli* deformylase	Peptide deformylase	100.0%	29	2w3t
Crystal structure of peptide deformylase from *Ehrlichia chaffeensis*	Peptide deformylase	100.0%	28	3oca
Crystal Structure of *P. aeruginosa* Peptide deformylase complexed with antibiotic Actinonin	Peptide deformylase	100.0%	31	1ix1
High resolution crystals structure of Cobalt-peptide deformylase bound to formate	Peptide deformylase	100.0%	29	1xeo

Table 4 | A *H. hamelinensis* contig shows predicted structural similarity to peptide deformylase. The 5′ to 3′ frame 3 translated of *H. hamelinensis* genome contig AJRK01000282 was used as a query sequence for Phyre2 structural prediction. The top four results are shown. This sequence is modeled with 100% confidence to known peptide deformylase structures.

Hit Name	Hit Molecule	Confidence	% ID	PDB
Crystal structure of peptide deformylase from *Ehrlichia chaffeensis*	Peptide deformylase	100.0%	42	3oca
Crystal Structure of *P. aeruginosa* Peptide deformylase complexed with antibiotic Actinonin	Peptide deformylase	100.0%	41	1ix1
Crystals of Peptide Deformylase from *Plasmodium falciparum* with Ten Subunits per Asymmetric Unit Reveal Critical Characteristics of the Active Site for Drug Design	Peptide deformylase	100.0%	32	1jym
Plant peptide deformylase PDF1B crystal structure	Peptide deformylase	100.0%	40	3cpm

Table 5 | *Methanocorpusculum bavaricum* genome shows synteny with *def-fmt* region of *M. labreanum* genome. A BLAST search against the nr database was performed using *M. bavaricum* ORFs upstream and downstream of putative *def-fmt* gene pair as query sequences. The top two hits are shown for each ORF, where P: indicates the protein name, S: indicates the species and A: indicates the accession number. All of the *M. bavaricum* ORFs that are visually equivalent to *M. labreanum* ORFs shows top hits to *M. labreanum* (highlighted in yellow).

Sequence Name	Top BLAST Hit	
	Annotation	E-value
Mbar_1 (position equivalent to Mlab_188)	**P:** Adenylyltransferase **S:** *Methanocorpusculum labreanum* **A:** WP_011832565.1	$2x10^{-153}$
	P: Adenylyltransferase **S:** *Methanolobus psychrophilus* **A:** WP_015052949.1	$2x10^{-83}$
Mbar_2 (position equivalent to Mlab_189)	**P:** Hypothetical protein **S:** *Methanocorpusculum labreanum* **A:** WP_011832566.1	$1x10^{-49}$
	P: Molybdopterin synthase subunit MoaD **S:** *Methanoplanus limicola* **A:** WP_004078578.1	$3x10^{-37}$
Mbar_3 (position equivalent to Mlab_190)	**P:** Hypothetical protein **S:** *Methanocorpusculum labreanum* **A:** WP_011832567.1	$1x10^{-33}$
	P: Hypothetical protein **S:** *Methanoculleus marisnigri* **A:** WP_011845255.1	$1x10^{-23}$
Mbar_4 (position equivalent to Mlab_191)	**P:** Peptide deformylase **S:** *Methanocorpusculum labreanum* **A:** WP_011832568.1	$4x10^{-102}$
	P: Peptide deformylase **S:** *Sebaldella termitidis* **A:** WP_012860638.1	$1x10^{-52}$
Mbar_5 (position equivalent to Mlab_192)	**P:** Methionyl-tRNA formyltransferase **S:** *Methanocorpusculum labreanum* **A:** WP_011832569.1	0.0
	P: Methionyl-tRNA formyltransferase **S:** *Ilyobacter polytropus* **A:** WP_013387131.1	$2x10^{-105}$
Mbar_6 (position equivalent to Mlab_193)	**P:** Hypothetical protein **S:** *Methanocorpusculum labreanum* **A:** WP_011832570.1	0.0
	P: Multidrug ABC transporter substrate-binding protein **S:** *Methanoculleus marisnigri* **A:** WP_011844657.1	$1x10^{-84}$
Mbar_7 (position equivalent to Mlab_194)	**P:** Macrolide ABC transporter ATP-binding protein **S:** *Methanocorpusculum labreanum* **A:** WP_011832571.1	$8x10^{-158}$
	P: Macrolide ABC transporter ATP-binding protein **S:** *Methanofollis liminatans* **A:** WP_004039285.1	$3x10^{-107}$
Mbar_8 (position equivalent to Mlab_195)	**P:** Hypothetical protein **S:** *Methanocorpusculum labreanum* **A:** WP_011832572.1	0.0
	P: Hypothetical protein **S:** *Methanofollis liminatans* **A:** WP_004039284.1	$7x10^{-57}$

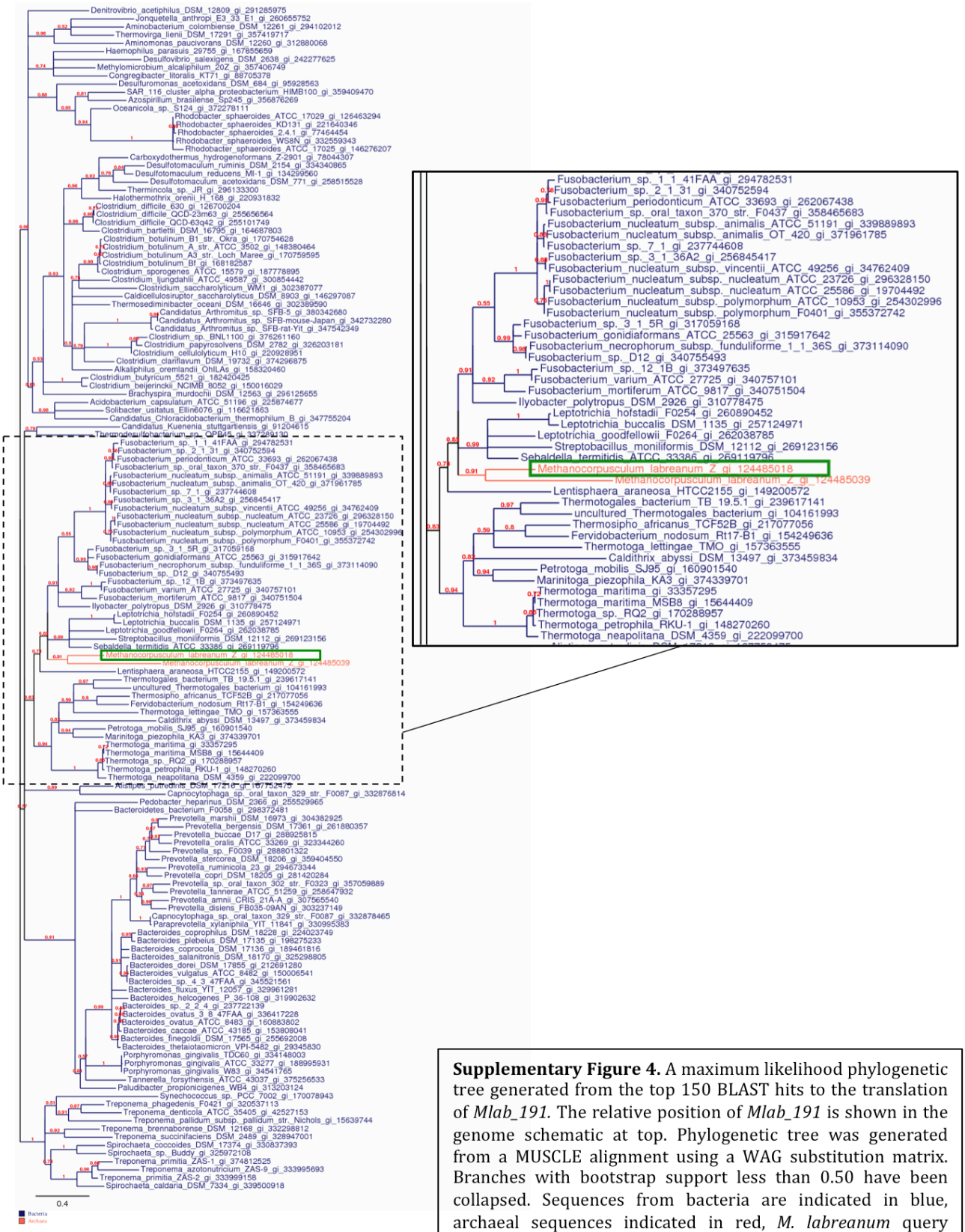
Table 6 | *H. hamelinensis* genome contig AJRK01000282 encodes peptide deformylase. A six-frame translation of the 395bp contig AJRK01000282 shows two frames that are not interrupted by stop codons. One of these (highlighted in yellow) shows strong similarity (E-value $< 10^{-80}$) to known peptide deformylase enzymes from bacterial species.

Translation	Sequence	Top BLAST hits		Top pHMMER hits	
		Name	**E-value**	**Name**	**E-value**
5′ to 3′ Frame 1	PMTCWKPCTLRRVLGLLRRRSGCWIG*SC SIASRKRTVRRARWSCSTPKSWRPRTRRT SMRKGACRFPTNTPR*PAPRWSRSNGWI VMAMLSGRHSTGFGPPVSSMRSTI*MASC S*TT*NR*SGR*SPLR	Hypothetical Protein *Catenovulum agarivorans* WP_016954737.1	1.2	Uncharacterized Protein *Strongylocentrotus purpuratus* H3JL95_STRPU	3.3
		Hypothetical Protein *Streptococcus sanguinis* WP_002898530.1	1.7	Uncharacterized Protein (Precursor) *Pseudomonas sp.* GM60 J3BB21_9PSED	6.5
5′ to 3′ Frame 2	R*HVGNHVRCAGYWACCAADRGAGSAD RARLRQGRGRSGAPVGHVQPRSRGVLGR GERL*GRVPVDSRPIRRGDPPQGGRGRMD GS*WQCSAGDIRRALGHLCPA*DRPSEWQ AVHRLPETAEAADDHP*D	No hits	-	No hits	-
5′ to 3′ Frame 3	DDMLETMYAAPGIGLAAPQIGVLDRLIVL DCVKEEDGPARPLVMFNPEVVASSDEAN VYEEGCLSIPDQYAEVTRPKVVEVEWMD RDGNAQRETFDGLWATCVQHEIDHLNGK LFIDYLKPLKRQMITPET	Peptide Deformylase *Phaeobacter inhibens* DSM 17395 YP_006571552.1	$2x10^{-88}$	Peptide Deformylase *Phaeobacter gallaeciensis* (strain 2.10) I7EQM3_PHAG2	$1.7x10^{-82}$
		Peptide Deformylase *Phaeobacter gallaeciensis* DSM 26640 YP_008974256.1	$2x10^{-88}$	Peptide Deformylase *Phaeobacter inhibens* (strain ATCC 700781) I7EIA5_PHAIB	$1.7x10^{-82}$
3′ to 5′ Frame 1	SLRGDHLPLQRFQVVYEQLAIQMVDLMLD TGGPKPVECLPLSIAITIHPFDLDHLGAGH LGVLVGNRQAPFLIDVRLVRGRHDFGVEH DQRARRTVLFLDAIEHDQPIQHPDLRRSK PNTRRSVHGFQHVI	Class I and II Aminotransferase *Nitratireductor indicus* WP_009451586.1	$1x10^{-18}$	Class I and II Aminotransferase *Nitratireductor indicus* C115 K2P1G2_9RHIZ	$1.2x10^{-18}$
		Class I and II Aminotransferase *Brucella canis* ATCC 23365 YP_001594969.1	$4x10^{-10}$	Aminotransferase, Class I and II *Brucella canis* (strain ATCC 23365) A9MCW1_BRUC2	$3.6x10^{-10}$
3′ to 5′ Frame 2	VSGVIICRFSGFR*SMNSLPFRWSISCWTQ VAQSPSNVSR*ALPSRSIHSTSTTLGRVTSA YWSGIDRHPSS*TFASSEDATTSGLNMTN GRAGPSSSLTQSSTISRSSTPICGAASPIPGA AYMVSNMSS	Hypothetical Protein *Brevibacillus sp.* CF112 WP_007786170.1	0.032	Ligase *Streptomyces griseoaurantiacus* M045 F3NTQ5_9ACTO	0.49
		Ligase Streptomyces griseoaurantiacus WP_006144239.1	0.076	Uncharacterized Protein (Precursor) *Brevibacillus sp.* CF112 J3AXZ4_9BACL	1
3′ to 5′ Frame 3	SQG*SSAASAVSGSL*TACHSDGRSHAGHR WPKARRMSPAEHCHHDPSIRPRPPWGGS PRRIGRESTGTLPHRRSPRPRTPRLRG*T* PTGAPDRPLP*RNRARSADPAPRSAAQQA QYPAQRTWFPTCHR	Chromodomain-Helicase-DNA-Binding Protein 1 *Rattus norvegicus* NP_001100935.1	5.5	*Rhodospirillum photometricum* DSM 122 H6SNS2_RHOPH	4
				Putative Cytochrome P450 *Ixodes ricinus* V5GPZ0_IXORI	8.4

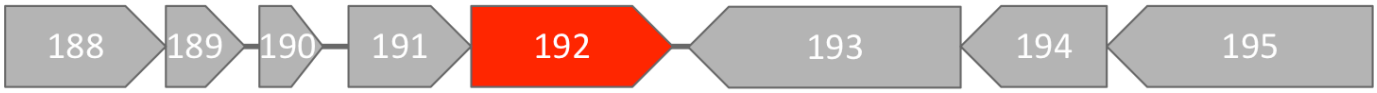
Supplementary Figure 1. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_188.* The relative position of *Mlab_188* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.

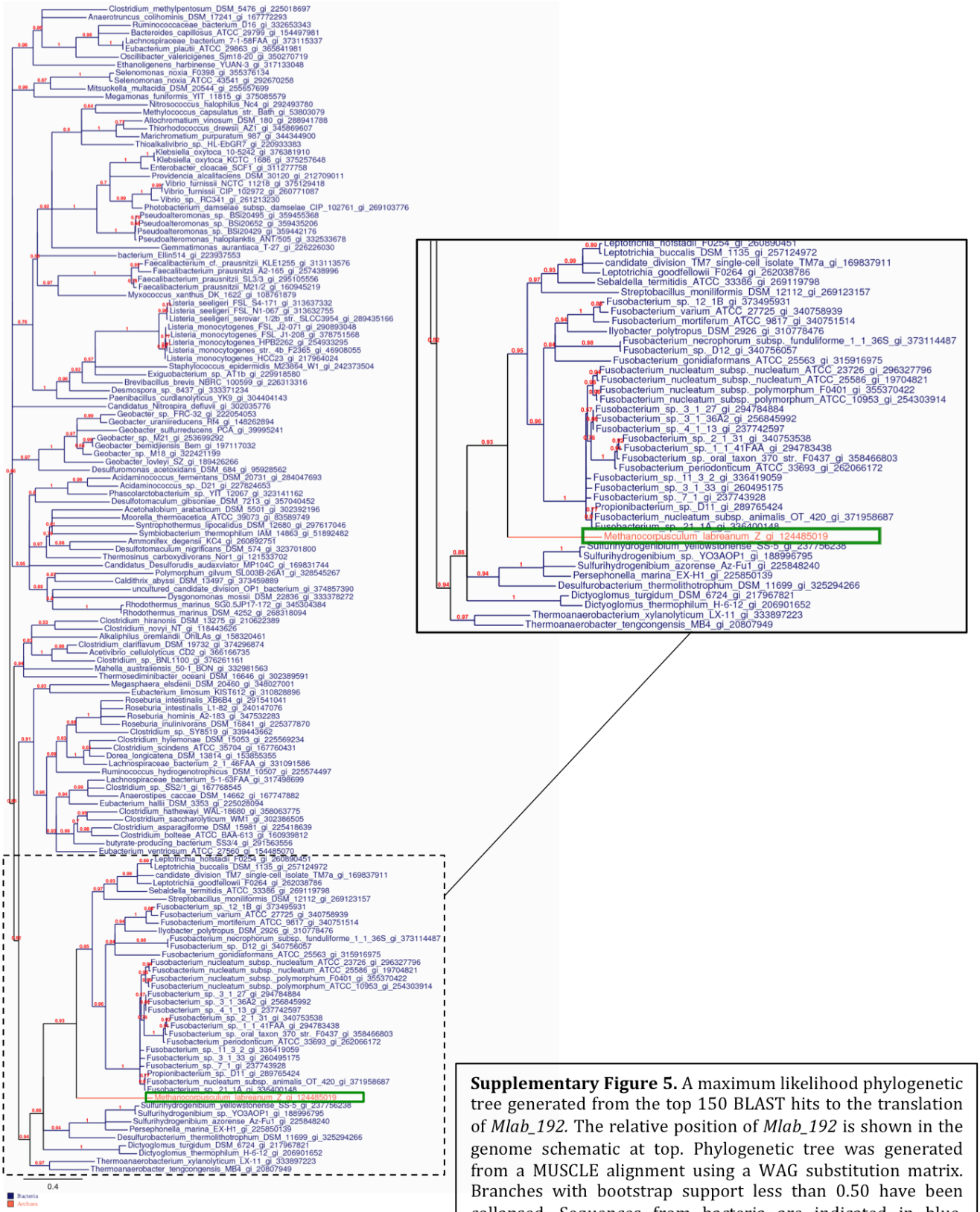
Supplementary Figure 2. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_189*. The relative position of *Mlab_189* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.

Supplementary Figure 3. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_190.* The relative position of *Mlab_190* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.

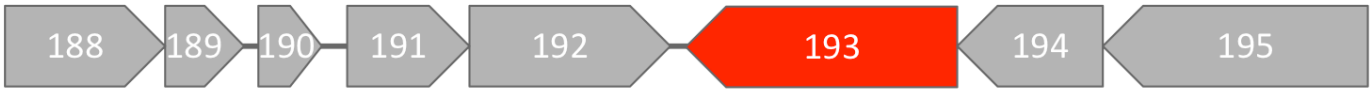


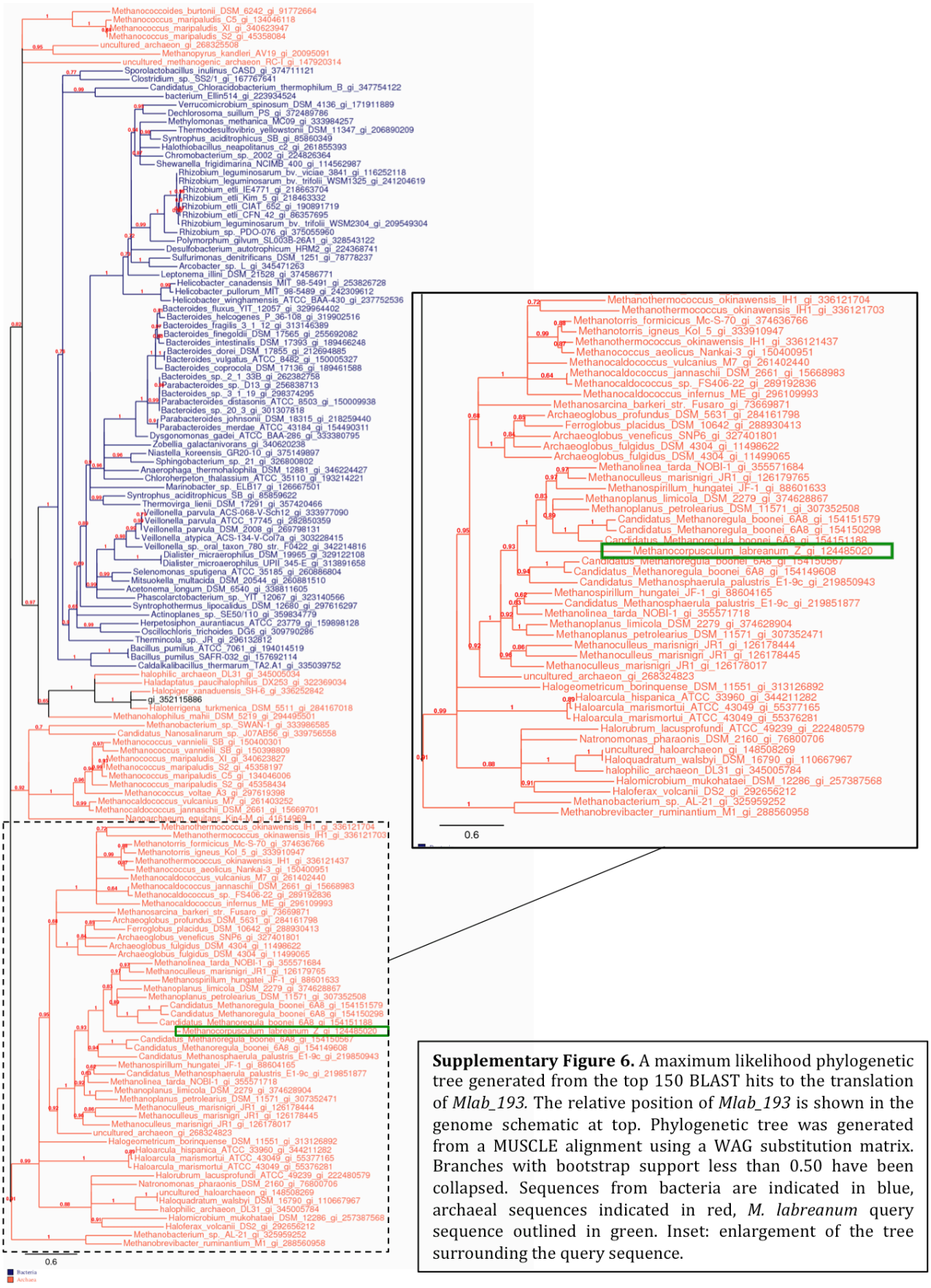
Supplementary Figure 4. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_191*. The relative position of *Mlab_191* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.





Supplementary Figure 5. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_192*. The relative position of *Mlab_192* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.





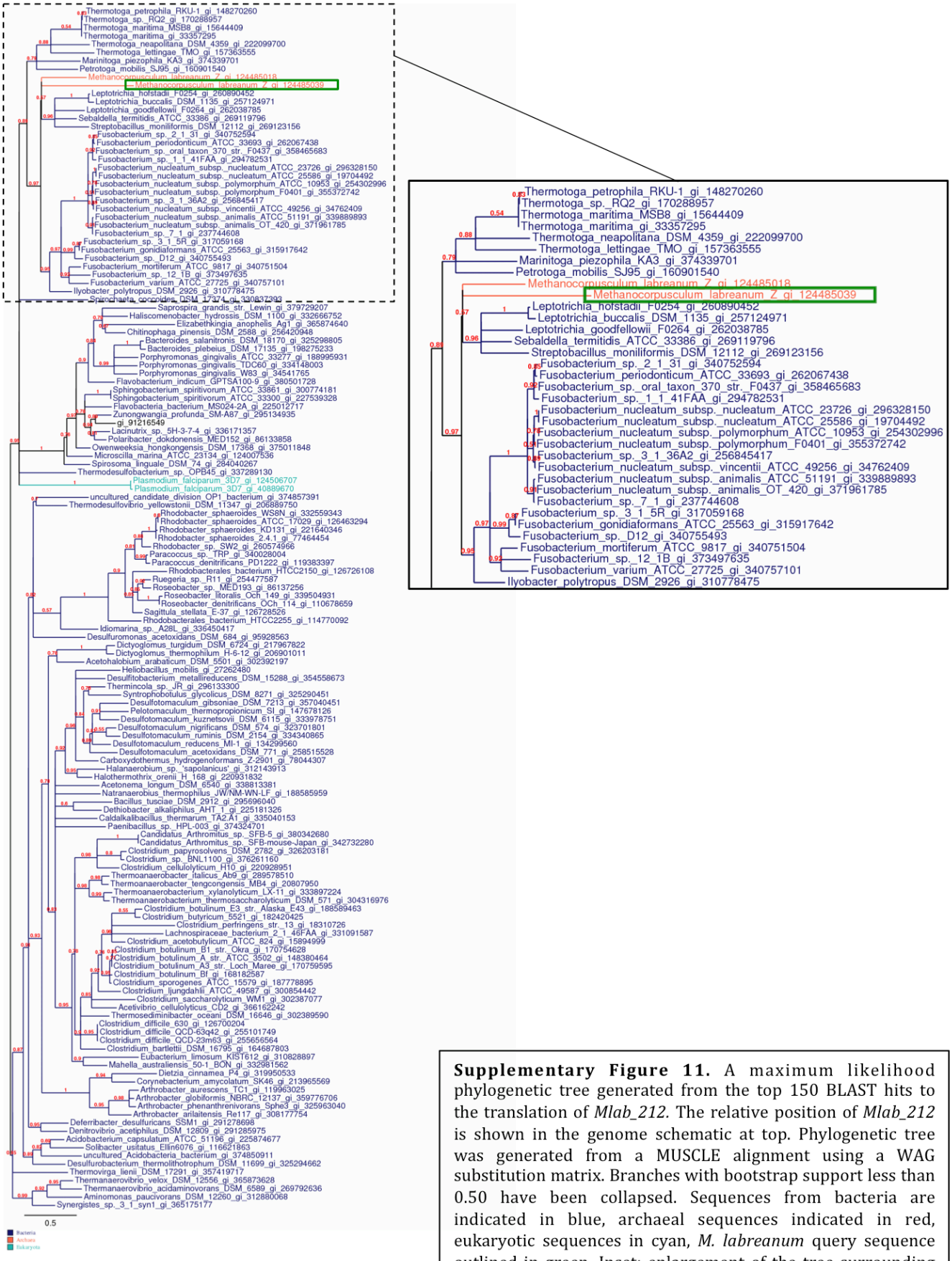
Supplementary Figure 6. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_193*. The relative position of *Mlab_193* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.

Supplementary Figure 7. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_194.* The relative position of *Mlab_194* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.

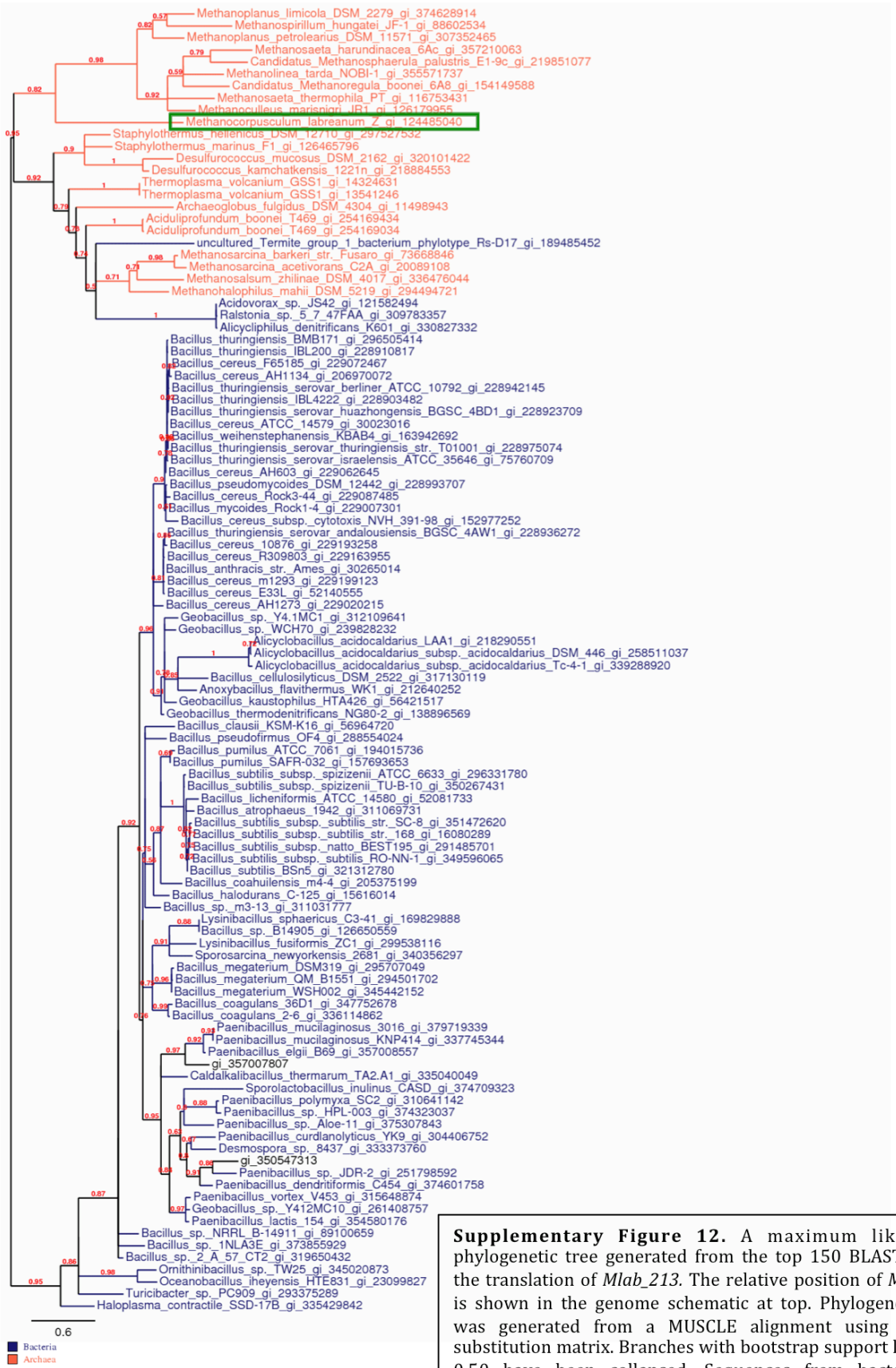
Supplementary Figure 8. A maximum likelihood phylogenetic tree generated from the BLAST hits with an E-value below 0.01 to the translation of *Mlab_195.* The relative position of *Mlab_195* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from environmental samples are indicated in blue, euryarchaeal sequences indicated in red, *M. labreanum* query sequence outlined in green.

Supplementary Figure 9. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_210.* The relative position of *Mlab_210* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.

Supplementary Figure 10. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_211*. The relative position of *Mlab_211* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, eukaryotic sequences in cyan, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.



Supplementary Figure 11. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_212*. The relative position of *Mlab_212* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, eukaryotic sequences in cyan, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.



Supplementary Figure 12. A maximum likelihood phylogenetic tree generated from the top 150 BLAST hits to the translation of *Mlab_213*. The relative position of *Mlab_214* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green. Inset: enlargement of the tree surrounding the query sequence.

Supplementary Figure 13. A maximum likelihood phylogenetic tree generated from the BLAST hits with an E-value below 0.01 to the translation of *Mlab_214*. The relative position of *Mlab_214* is shown in the genome schematic at top. Phylogenetic tree was generated from a MUSCLE alignment using a WAG substitution matrix. Branches with bootstrap support less than 0.50 have been collapsed. Sequences from bacteria are indicated in blue, archaeal sequences indicated in red, *M. labreanum* query sequence outlined in green.

References

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics *21*, 2104–2105.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Anderson, I.J., Sieprawska-Lupa, M., Goltsman, E., Lapidus, A., Copeland, A., Glavina Del Rio, T., Tice, H., Dalin, E., Barry, K., Pitluck, S., et al. (2009). Complete genome sequence of Methanocorpusculum labreanum type strain Z. Stand. Genomic Sci. *1*, 197–203.

Ando, S., Ishikawa, K., Ishida, H., Kawarabayasi, Y., Kikuchi, H., and Kosugi, Y. (1999). Thermostable aminopeptidase from Pyrococcus horikoshii. FEBS Letters *447*, 25–28.

Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., and Koonin, E.V. (1998). Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. Trends Genet. *14*, 442–444.

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolić, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B., and Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature *452*, 108–111.

Brochier-Armanet, C., and Forterre, P. (2007). Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. Archaea *2*, 83–93.

Burns, B.P., Gudhka, R.K., and Neilan, B.A. (2012). Genome sequence of the halophilic archaeon Halococcus hamelinensis. J. Bacteriol. *194*, 2100–2101.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552.

Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE *5*, e11147.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.

Garcia-Vallvé, S., Romeu, A., and Palau, J. (2000). Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. Genome Res. *10*, 1719–1725.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology *59*, 307–321.

Heijne, von, G. (1984). Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. EMBO J. *3*, 2315.

Heinemann, J.A., and Kurenbach, B. (2009). Horizontal transfer of genes between microorganisms. In Encyclopedia of Microbiology, (Elsevier Inc.).

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754–755.

Jain, R., Rivera, M.C., and Lake, J.A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. PNAS *96*, 3801–3806.

Jia, B., Lee, S., Pham, B.P., Kwack, J.M., Jin, H., Li, J., Wang, Y., and Cheong, G.-W. (2011). Biochemical characterization of deblocking aminopeptidases from the hyperthermophilic archaeon Thermococcus kodakarensis KOD1. Biosci. Biotechnol. Biochem. *75*, 1160–1166.

Julián, P., Milón, P., Agirrezabala, X., Lasso, G., Gil, D., Rodnina, M.V., and Valle, M. (2011). The Cryo-EM structure of a complete 30S translation initiation complex from Escherichia coli. PLoS Biol. *9*, e1001095.

Katoh, K., Misawa, K., Kuma, K.I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc *4*, 363–371.

Kozak, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. *47*, 1–45.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Lee, C.P., Seong, B.L., and RajBhandary, U.L. (1991). Structural and sequence

elements important for recognition of Escherichia coli formylmethionine tRNA by methionyl-tRNA transformylase are clustered in the acceptor stem. J. Biol. Chem. *266*, 18012–18017.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. *25*, 0955–0964.

Martens, T., Heidorn, T., Pukall, R., Simon, M., Tindall, B.J., and Brinkhoff, T. (2006). Reclassification of Roseobacter gallaeciensis Ruiz-Ponte et al. 1998 as Phaeobacter gallaeciensis gen. nov., comb. nov., description of Phaeobacter inhibens sp. nov., reclassification of Ruegeria algicola (Lafay et al. 1995) Uchino et al. 1999 as Marinovum algicola gen. nov., comb. nov., and emended descriptions of the genera Roseobacter, Ruegeria and Leisingera. Int. J. Syst. Evol. Microbiol. *56*, 1293–1304.

Meinnel, T., Guillon, J.M., Mechulam, Y., and Blanquet, S. (1993). The Escherichia coli fmt gene, encoding methionyl-tRNA(fMet) formyltransferase, escapes metabolic control. J. Bacteriol. *175*, 993–1000.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., and Dodson, R.J. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. Nature *399,* 323-329.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302,* 205-217.

Onoe, S., Ando, S., Ataka, M., and Ishikawa, K. (2002). Active Site of Deblocking Aminopeptidase from Pyrococcus horikoshii. Biochemical and Biophysical Research Communications *290*, 994–997.

Ramesh, V., and RajBhandary, U.L. (2001). Importance of the anticodon sequence in the aminoacylation of tRNAs by methionyl-tRNA synthetase and by valyl-tRNA synthetase in an Archaebacterium. J. Biol. Chem. *276*, 3660–3665.

Ruiz-Ponte, C., Cilia, V., Lambert, C., and Nicolas, J.L. (1998). Roseobacter gallaeciensis sp. nov., a new marine bacterium isolated from rearings and collectors of the scallop Pecten maximus. Int. J. Syst. Evol. Microbiol. *48*, 537–542.

Simonetti, A., Marzi, S., Myasnikov, A.G., Fabbretti, A., Yusupov, M., Gualerzi, C.O., and Klaholz, B.P. (2008). Structure of the 30S translation initiation complex. Nature *455*, 416–420.

Story, S.V., Grunden, A.M., and Adams, M.W.W. (2001). Characterization of an Aminoacylase from the Hyperthermophilic Archaeon Pyrococcus furiosus. J.

Bacteriol. *183*, 4259–4268.

Ushida, C., Muramatsu, T., Mizushima, H., Ueda, T., Watanabe, K., Stetter, K.O., Crain, P.F., McCloskey, J.A., and Kuchino, Y. (1996). Structural feature of the initiator tRNA gene from Pyrodictium occultum and the thermal stability of its gene product, tRNA(imet). Biochimie *78*, 847–855.

White, B.N., and Bayley, S.T. (1972a). Methionine transfer RNAs from the extreme halophile, Halobacterium cutirubrum. Biochim. Biophys. Acta *272*, 583–587.

White, B.N., and Bayley, S.T. (1972b). Methionine transfer RNAs from the extreme halophile, Halobacterium cutirubrum. Biochimica Et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis *272*, 583–587.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R.U.D., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J., et al. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature *462*, 1056–1060.

Yang, Z., Nielsen, R., and Hasegawa, M. (1998). Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. *15*, 1600–1611.

Zellner, G., Stackebrandt, E., Messner, P., Tindall, B.J., de Macario, E.C., Kneifel, H., Sleytr, U.B., and Winter, J. (1989). Methanocorpusculaceae fam. nov., represented by Methanocorpusculum parvum, Methanocorpusculum sinense spec. nov. and Methanocorpusculum bavaricum spec. nov. Arch. Microbiol. *151*, 381–390.

Zhao, Y., Boone, D.R., Mah, R.A., Boone, J.E., and Xun, L. (1989). Isolation and Characterization of Methanocorpusculum labreanum sp. nov. from the LaBrea Tar Pits. Int J Syst Bacteriol *39*, 10–13.

Zheng, Y., Roberts, R.J., Kasif, S., and Guan, C. (2005). Characterization of Two New Aminopeptidases in Escherichia coli. J. Bacteriol. *187*, 3671–3677.

Chapter 7
Discussion

The use of formylation of the initiating methionine in bacterial translation was discovered 50 years ago by Kjeld Marcker and Frederick Sanger (Marcker and Sanger, 1964). In the half-century since its discovery, many hypotheses have been proposed for the function of formylation. However none of these hypotheses satisfactorily explain why the requirement for formylation is restricted to the bacterial domain nor do they explain the great variation in dependence upon deformylation. The conservation of the ribosome, core initiation factors and initiator tRNA suggest that any requirement for formylation would be shared across all domains of life. The results presented in this body of work suggest a non-adaptive role for formylation in bacteria. The findings in Chapter 2 show unambiguously that Met-tRNA$_i^{fMet}$ formylation is not required for translation in *E. coli,* in agreement with previous reports (Guillon et al., 1992; Mazel et al., 1994). Chapter 3 shows that formylation is not required for wild-type growth rates in *E. coli.* Therefore, the ubiquity of formylation in bacteria is not reconcilable with an adaptive function. Our results also show that multiple mutations are required to compensate for the loss of formylation. Chapter 4 shows that reintroduction of the formylation-deformylation cycle to our growth-compensated, non-formylating *E. coli* produces a phenotype that is not consistent with complementation. Rather, a *def-fmt* encoding plasmid causes post-segregational killing upon loss from DF200. In chapter 5, we show that the toxic phenotype of formylation that is a requirement for PSK in *E. coli* is not lethal in *S. cerevisiae,* nullifying the addictive phenotype of formylation. Finally, in chapter 6 we show evidence for the horizontal transfer of formylation from bacteria to archaea. These final observations are consistent with the hypothesis that formylation and deformylation

initially evolved as part of a competition strategy between horizontally mobile genetic elements that were endemic to bacteria. These genes migrated to the chromosome, as do many toxin-antitoxin pairs, as part of how bacteria adapted to the fitness costs associated with the competition.

Model for evolution of formylation in bacteria

Based on these results, we propose a model for the evolution of formylmethionine use in bacterial translation (Figure 1). In our model, the ancestors of bacteria, archaea and eukarytoes did not utilize Met-tRNA$_i$ formylation, initiating translation with methionine. The acquisition of a methionyl-tRNA formyltransferase gene and peptide deformylase gene by a mobile genetic element, such as a plasmid, conferred upon it PSK activity. As ancestral cells initiated translation with methionine, the introduction of a formylase gene on a plasmid and the resulting formylation of Met-tRNA$_i$ would likely present a hindrance to translation. We expect the ancestral initiation factors, especially IF2, would have lower affinity for the foreign, uncharged fMet-tRNA$_i$ than the native, positively charged Met-tRNA$_i$, slowing translation initiation. Despite this, a formylase-deformylase encoding plasmid could spread through an ancestral population by horizontal gene transfer. Additionally, the necessity for peptide deformylation in the presence of fMet-tRNA$_i$ would ensure the plasmid was successful at competing with other plasmids. This is because when two incompatible plasmids occupy the same cell through HGT, they segregate into different daughters. The daughter not inheriting the deformylase gene, and the plasmid it carries, would die. This competition imposes a significant fitness cost to bacteria. There is selection for cells that capture the gene pair on the chromosome, and which are then immunised from the competition (Cooper and Heinemann, 2000; 2005; Saavedra De Bast et al., 2008).

As is documented for modern toxin-antitoxin pairs that elicit PSK, rare individuals in the population may escape toxin activity by inactivation of the formylase-encoding gene (Cooper and Heinemann, 2000; Naito et al., 1995). However, following subsequent loss also of the deformylase-encoding gene, individuals become susceptible again to hosting competition between mobile genetic elements with and without the *def-fmt* gene pair. In this way, a toxin-antitoxin pair that also causes PSK is able to spread to fixation in a population, despite its susceptibility to loss by ordered gene inactivation(Rankin et al., 2012). Due to the spread of plasmids and persistent infection/reinfection of hosts by plasmids, cells would spend more time with formylase activity than in the absence of it. Consequently, when plasmids with *def-fmt* genes are common, any cells that adapt to the presence of fMet-$tRNA_i$, initiating translation more efficiently, would have a fitness advantage. We predict that these adaptations to the presence of formylation would impact a similar set of genes to those observed to mutate in our long-term evolution experiment (see Chapter 3). As the population adapts to the presence of fMet-$tRNA_i$, the loss of formylation becomes disadvantageous, similar to modern bacteria (Chapter 1, Table 1). In this state, the loss of deformylase activity continues to be lethal due to the accumulation of formylated proteins, however, unlike canonical toxin-antitoxin pairs, the loss of formylase activity (our 'toxin' equivalent) is also detrimental, resulting in a large growth disadvantage. In effect, cells are now unable to escape a fitness penalty from the loss of either the formylase or the deformylase-encoding gene.

Our long-term culturing experiment resulted in the emergence of a non-formylating bacterial strain with a wild-type growth rate. In our model, this represents the ancestral state of translation without formylation. It is worth noting that, although our DF200 lines show features consistent with the putative ancestral bacteria described by our model (efficient growth in the absence of formylation), they are only an approximation of an ancestral genotype. It is impossible to know the full

nature of ancestral bacteria and the evolutionary trajectory that facilitated adaptation from a non-formylating to a formylating state may not be reversible. Moreover, there may be multiple genotypes that facilitate wild-type growth rates in the absence of formylation. Our DF200 lines had a mutation in the promoter for *metZ*, amplifying expression of the initiator tRNA (Chapter 3, Figure 6). Such amplification has been shown previously to facilitate translation despite a low affinity of Met-$tRNA_i^{fMet}$ for IF2 (Nilsson et al., 2006). Additionally, mutations in *infB* that increase the affinity of IF2 for Met-$tRNA_i^{fMet}$ can also facilitate translation in the absence of formylation (Steiner-Mosonyi et al., 2004; Zorzet et al., 2010). Thus, there are at least two different mechanisms by which cells can overcome the low binding affinity of IF2 for non-formylated Met-$tRNA_i^{fMet}$. Incidentally, this Met-$tRNA_i^{fMet}$ amplification may also mitigate the predicted growth disadvantage caused by re-introduction of formylation to DF200 strains. The increased Met-$tRNA_i^{fMet}$ is thought to overcome the low affinity of IF2 for non-formylated Met-$tRNA_i^{fMet}$ by providing additional substrate for the binding reaction (Nilsson et al., 2006). In the same way, this additional initiator tRNA may allow binding of fMet-$tRNA_i^{fMet}$ to IF2 in our DF200 strain, mitigating any growth disadvantage.

Model for the maintenance of formylation in bacteria

Extrapolating from our model, we can hypothesise on the mechanism by which the formylation-deformylation cycle is maintained in bacteria. As observed in chapter 4, loss of the formylase-deformylase gene pair carried by a mobile genetic element would result in PSK. Immunity to this effect would follow from acquisition of the gene pair to the chromosome. Further loss of the toxin – formylation activity - would have to proceed through ordered formylase-deformylase gene loss. We theorise that the transition from a formylating state (as in modern bacteria) to a non-formylating state (as in our evolved knockout strains) is not hindered by the

loss of an important function of formylation in the cell. Rather, this transition requires that cells shift from a state where the cell biochemistry is adapted to initiation with fMet-tRNA$_i$, to one adapted to initiation with Met-tRNA$_i$. The intermediate state in this transition is one where cells are adapted to initiation with fMet-tRNA$_i$, but no longer utilize formylation, therefore must initiate translation with Met-tRNA$_i$. This maladapted state is similar to our unevolved knockout (Chapter 2), and the formylase knockout strains identified previously (Chapter 1, Table 1). This fMet-tRNA$_i$ to Met-tRNA$_i$ transition can be represented by a fitness landscape (Figure 2) where both the fMet-adapted and Met-adapted conditions are states of high fitness separated by a valley of low fitness (representing intermediate states). In the laboratory conditions we created, low fitness strains could reproduce without competition from modern, more fit genotypes and give rise incrementally to better adapted offspring during our long-term evolution experiment. *E. coli* required at least 800 generations to traverse this valley of low fitness. In a natural competitive environment, we expect cells that fall in to this valley of low fitness would be quickly outcompeted and lost from the population.

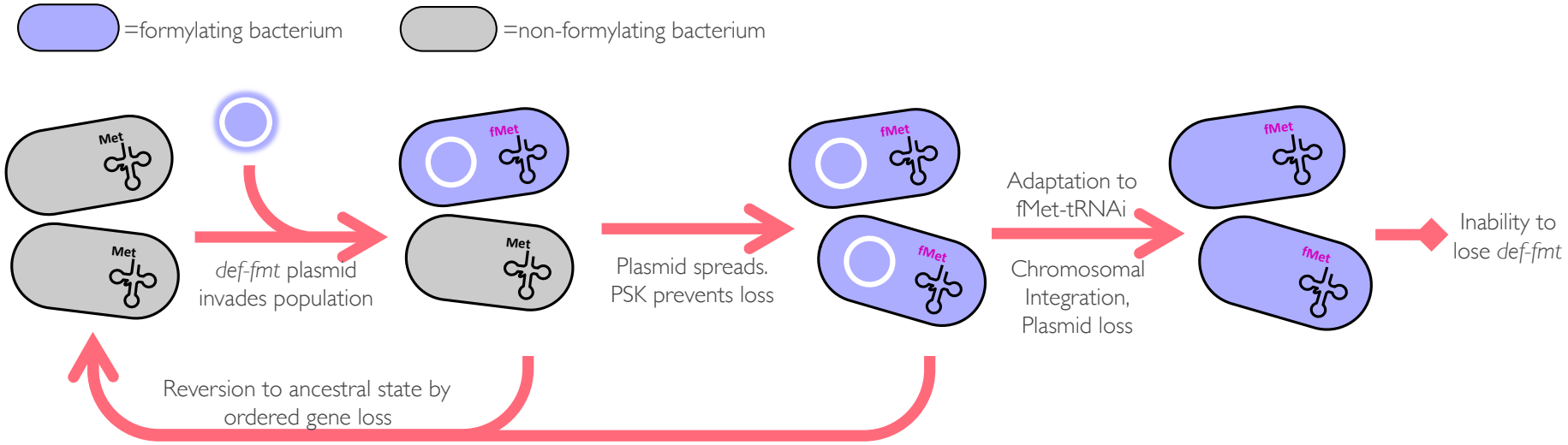


Figure 1 | Model for the evolution of formylation in bacteria. The ancestors of bacteria initiated translation with non-formylated methionine (grey cells). This provided a niche for occupancy by a formylase-deformylase encoding plasmid. The presence of such a plasmid would result in formylation of Met-tRNA$_i$, introducing a dependency on peptide deformylation for survival (purple cells). This dependency is the basis of the PSK phenotype elicited by deformylase-formylase encoding plasmids in non-formylating bacteria. The horizontally mobile plasmid would spread through the population, transforming non-formylating cells to a formylating phenotype and ensuring its maintenance in populations by PSK. Escape from PSK and reversion of cells to an ancestral state would make individuals susceptible to plasmid re-infection, thus ensuring populations are in a formylating state more often than non-formylating. This continual formylase and deformylase presence would select for individuals that adapt to the presence of fMet-tRNA$_i$ through compensatory mutation. Integration of the formylase-deformylase encoding plasmid would prevent PSK, resulting in plasmid loss. As the population adapts to a formylation-proficient state, the loss of formylation becomes detrimental.

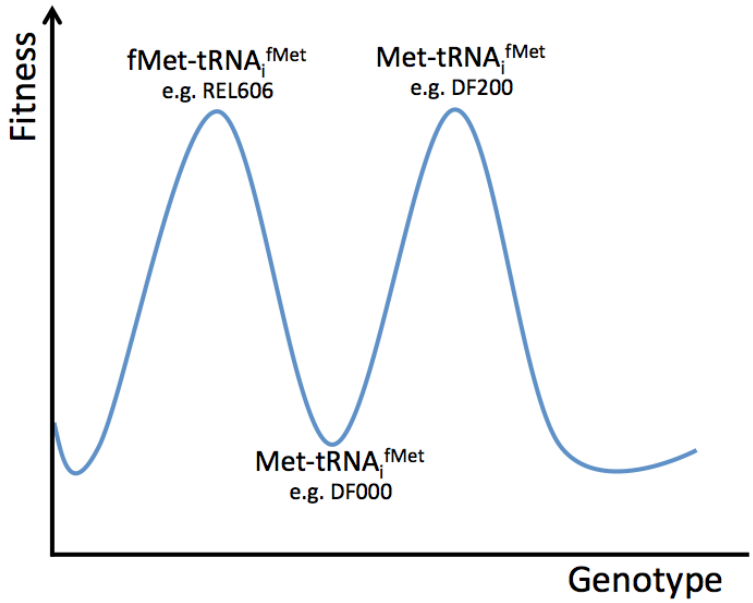


Figure 2 | Fitness landscape of fMet-to-Met transition. The loss of fMet-tRNA$_i^{fMet}$ in bacterial translation can be visualized with a fitness landscape. Adaptation to fMet-tRNA$_i^{fMet}$ use (as in wild-type bacteria e.g. REL606) represents a local fitness maximum, as does adaptation to Met-tRNA$_i^{fMet}$ (as in our long-term evolution experiment e.g. DF200). In contrast, non-formylating strains without adaptation to Met-tRNA$_i^{fMet}$ represent an intermediate valley of low fitness.

Formylation can be lost from bacteria

The capacity of *E. coli* to adapt to a non-formylating state through little more than growing non-formylating strains in isolation from faster growing wild-type strains suggests a further hypothesis based on our model. We predict that spontaneously arising formylation-deficient cells may not be outcompeted if they are isolated from formylation-proficient populations for a sufficient period of time. Such a situation might be possible under conditions of frequent bottle-necking

(where population contraction serendipitously removes formylating competitors), or conditions where selection is minimized and genetic drift is a dominant force on population dynamics (allowing non-formylating cells to increase in frequency despite low fitness). Indeed, during the preparation of this thesis, data was published suggesting that the formylase-deformylase gene pair has been lost on three separate occasions during the evolution of Mollicutes (Grosjean et al., 2014). Specifically, the formylase-deformylase gene pair has been lost in two clades of *Mycoplasma sp.* and one clade of *Phytoplasma sp.* Species of *Mycoplasma* and *Phytoplasma* are specialized symbiotic or pathogenic bacteria, living within or on the surface of eukaryotic host cells. As such, we expect *Mycoplasma* and *Phytoplasma* populations to be subject to repeated genetic bottlenecks during transmission between hosts. Additionally, the intracellular environment may provide isolation from competing bacterial populations and limit population sizes of successful individuals, allowing the influence of genetic drift to prevail. Genetic drift has been shown to be a major contributor to the evolution of other obligate intracellular bacteria (Herbeck et al., 2003; Wernegreen and Moran, 1999), and thus is likely to be influential in *Mycoplasma* and *Phytoplasma* genome evolution. Consequently, the loss of formylation from these species may reflect evolution in an environment favouring drift over competition and selection.

This ability of bacteria to overcome the requirement for formylation through isolation from formylation-proficient competitors may have important ramifications for the success of antibiotics. Peptide deformylase inhibitors are being developed as antibacterial agents (Clements et al., 2001; Giglione et al., 2000), with some being tested in clinical trials (O'Dwyer et al., 2013). The use of such antibiotics would impede the growth of formylation-proficient bacteria, allowing non-formylating bacteria to grow unchecked. Consequently, the use of peptide-deformylase inhibitors as antibiotics may provide an environment in which formylation-deficient bacteria similar to our DF200 lines could evolve.

Formylase evolution resembles constructive neutral evolution

Interestingly, our model of the evolution of formylation resembles theories of constructive neutral evolution. Constructive neutral evolution describes how complexity can arise without providing an adaptive advantage to hosts (Gray et al., 2010). In this theory, mutations that inactivate or decrease the efficiency of protein function can be compensated by subsequent mutations in interacting components. Such compensatory mutations are able to take multiple forms, whereas reversion to the native state requires a specific change, thus making compensatory mutation more likely. However, compensatory mutation, while restoring the native functionality of the system, takes the system further from the original state. Consequently, this progression acts as an evolutionary ratchet, favouring divergence from the native state over reversion (Gray et al., 2010). In this way, constructive neutral evolution increases the complexity of systems without conferring a selective advantage or increased functionality. In our formylation model, introduction of formylase and deformylase activities to ancestral cells results in an impairment in translation that can be relieved by eliminating formylation, or adapting to its presence. As explained earlier, elimination of formylation will result in re-establishment of the system following re-infection with a *def-fmt*-encoding plasmid. However, adaptation to the presence of fMet-tRNA$_i$ involves modification of the core components of translation to restore efficient translation initiation. If this adaptation were to take place, reversion to the ancestral state would require simultaneous elimination of formylation and alteration of translation components, with either reversion being detrimental alone. Each modification that compensates for the presence of formylation would take the system further from the ancestral state, reducing the likelihood of reversion. Through this mechanism, formylation becomes an integral part of translation initiation, despite providing no beneficial activity. In this way, the process of

constructive neutral evolution increases the complexity of translation, without providing any additional function.

Model for evolution of formylation in eukaryotes

Based on our results using both *E. coli* and *S. cerevisiae,* we have also developed a model to explain the absence of formylation from eukaryotic translation (Figure 3). The formylase-deformylase gene pair almost certainly entered eukaryotes through the bacterial ancestors of the mitochondria and plastids. Formylation and deformylation are still used in modern mitochondria and plastids, however the genes encoding these functions are not encoded in these organelles. Through a process of horizontal gene transfer (from free-living bacteria to ancestral eukaryotes) or endosymbiont gene transfer (from mitochondrial/plastid ancestors to eukaryotes), the formylase and deformylase genes were transferred to the nuclear genome. Similar to bacteria, the loss of the formylase and deformylase genes from mitochondrial and plastid ancestors would be lethal. Consequently, successful transfer of these genes from mitochondria and plastids to the nucleus would require targeting of formylase and deformylase proteins back to the organelles. The expression of formylase and deformylase genes from the nuclear genome may result in formylation of the cytoplasmic initiator tRNA (Met-$tRNA_i^{Met}$). However, in the presence of deformylase activity, this has a minor effect on eukaryotic growth. Additionally, formylation alone is not lethal in *S. cerevisiae,* abrogating any potential PSK activity. As a result, the loss of both formylation and deformylation from cytoplasmic translation would not be lethal to eukaryotes.

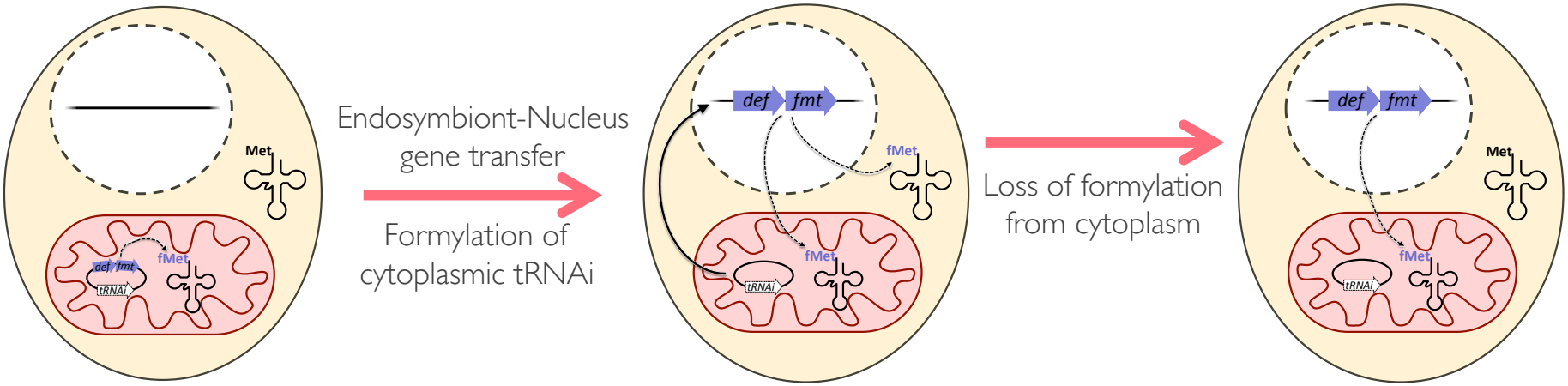


Figure 3 | Model for the evolution of formylation in eukaryotes. Genes encoding formylase and deformylase activity entered eukaryotes through the mitochondria and plastids. Due to their ancestral origin as bacterial cells, these organelles require formylation of Met-tRNA and peptide deformylation in translation. Our model predicts the movement of these genes from organelles (represented as a red mitochondrion) to the nucleus (represented as a white organelle, bordered by a dashed line) by endosymbiont-nucleus gene transfer. Due to the PSK activity of formylation and deformylation, this transfer to the nucleus would require targeting of these activities back to the organelles. Additionally, expression from a nuclear genome may result in formylation of the cytoplasmic Met-tRNA$_i^{Met}$. In the presence of deformylation, this has little effect on the growth of *S. cerevisiae*. Moreover, formylation without *def*-directed deformylation is not toxic in cytoplasmic translation. Consequently, the loss of formylase and deformylase activity from the cytoplasm would not result in PSK in eukaryotes, despite a difference in 'toxin' and 'antitoxin' lability. As a result, translation initiation in the cytoplasm can be restored to a non-formylating state, provided formylase and deformylase activity continues to be targeted to mitochondria and plastids.

Future directions

The data we have presented provide support for the nontrivial hypothesis that the spread of the *def-fmt* gene pair was through an ancient non-formylating bacterial population by PSK (see Chapter 4). Additionally, our genome sequencing data provides insights into the genes that may be responsible for adaptation to formylation (see Chapter 3). However, our long-term culturing experiment evolved a non-formylating bacterium from a formylation-proficient ancestor. In contrast, our model predicts the evolution of a formylation-proficient bacterium from a non-formylating ancestor. As mentioned above, there is no requirement for the evolutionary trajectory from formylating to non-formylating to be reversible. Consequently, an important facet of our model remains to be tested: how does a non-formylating bacterial population adapt to the presence of the *def-fmt* gene pair? This could easily be tested by re-introducing the *def-fmt* gene pair to DF200 and serially culturing the resulting strain.

The fitness of our DF200 line has not been measured. Although previous studies have shown that growth rate bears a strong correlation with fitness (Vasi et al., 1994), a competition experiment between DF200 and our evolved wild-type lines would provide a direct measurement of fitness. These data would provide insight into whether, given the chance to arise in the absence of competition, a non-formylating bacterium could persist in a formylation-proficient population. Such information may be important in determining whether deformylase-inhibitor-resistant bacteria could become a notable pathogen.

The competition model for the evolution of PSK posits that the selective advantage PSK confers on a plasmid is realised under conditions of within-host competition. That is, a horizontally mobile plasmid that elicits one unique manifestation of PSK

is able to outcompete an isogenic plasmid that lacks that one PSK system. We showed that the *def-fmt* gene pair is able to increase the stability of a plasmid in a monoculture of DF200, and induce PSK upon loss from DF200 (see Chapter 4). Although the *def-fmt* gene pair fulfils these diagnostic requirements of a PSK system, a direct test of the competitive ability of a *def-fmt*-encoding plasmid would provide further support of our model for the spread of *def-fmt* through ancestral populations.

We have shown that *def-fmt* homologues are found in archaea (see Chapter 6), though we have yet to provide biochemical evidence of their functionality in these *Methanocorpusculum sp.* A urea-PAGE separation followed by northern blotting analysis of the initiator tRNA from these species would provide direct evidence for formylation in archaea.

References

Clements, J.M., Beckett, R.P., Brown, A., Catlin, G., Lobell, M., Palan, S., Thomas, W., Whittaker, M., Wood, S., Salama, S., et al. (2001). Antibiotic activity and characterization of BB-3497, a novel peptide deformylase inhibitor. Antimicrob. Agents Chemother. *45*, 563–570.

Cooper, T.F., and Heinemann, J.A. (2000). Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. PNAS *97*, 12643–12648.

Cooper, T.F., and Heinemann, J.A. (2005). Selection for plasmid post-segregational killing depends on multiple infection: evidence for the selection of more virulent parasites through parasite-level competition. Proc. R. Soc. B *272*, 403–410.

Giglione, C., Pierre, M., and Meinnel, T. (2000). Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. Mol. Microbiol. *36*, 1197–1205.

Gray, M.W., Lukes, J., Archibald, J.M., and Keeling, P.J. (2010). Irremediable complexity? Science.

Grosjean, H., Breton, M., Sirand-Pugnet, P., Tardy, F., Thiaucourt, F., Citti, C., Barré, A., Yoshizawa, S., Fourmy, D., de Crécy-Lagard, V., et al. (2014). Predicting the minimal translation apparatus: lessons from the reductive evolution of mollicutes. PLoS Genet *10*, e1004363.

Guillon, J.M., Mechulam, Y., Schmitter, J.M., Blanquet, S., and Fayat, G. (1992). Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of Escherichia coli. J. Bacteriol. *174*, 4294–4301.

Herbeck, J.T., Funk, D.J., Degnan, P.H., and Wernegreen, J.J. (2003). A conservative test of genetic drift in the endosymbiotic bacterium Buchnera: slightly deleterious mutations in the chaperonin groEL. Genetics *165*, 1651–1660.

Marcker, K., and Sanger, F. (1964). N-Formyl-methionyl-S-RNA. J. Mol. Biol. *8*, 835–IN838.

Mazel, D., Pochet, S., and Marliere, P. (1994). Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. EMBO J. *13*, 914–923.

Naito, T., Kusano, K., and Kobayashi, I. (1995). Selfish behavior of restriction-

modification systems. Science *267*, 897–899.

Nilsson, A.I., Zorzet, A., Kanth, A., Dahlström, S., Berg, O.G., and Andersson, D.I. (2006). Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. PNAS *103*, 6976–6981.

O'Dwyer, K., Hackel, M., Hightower, S., Hoban, D., Bouchillon, S., Qin, D., Aubart, K., Zalacain, M., and Butler, D. (2013). Comparative analysis of the antibacterial activity of a novel peptide deformylase inhibitor, GSK1322322. Antimicrob. Agents Chemother. *57*, 2333–2342.

Rankin, D.J., Turner, L.A., Heinemann, J.A., and Brown, S.P. (2012). The coevolution of toxin and antitoxin genes drives the dynamics of bacterial addiction complexes and intragenomic conflict. Proc. R. Soc. B *279*, 3706–3715.

Saavedra De Bast, M., Mine, N., and Van Melderen, L. (2008). Chromosomal toxin-antitoxin systems may act as antiaddiction modules. J. Bacteriol. *190*, 4603–4609.

Steiner-Mosonyi, M., Creuzenet, C., Keates, R.A.B., Strub, B.R., and Mangroo, D. (2004). The Pseudomonas aeruginosa initiation factor IF-2 is responsible for formylation-independent protein initiation in P. aeruginosa. J. Biol. Chem. *279*, 52262–52269.

Vasi, F., Travisano, M., and Lenski, R.E. (1994). Long-term experimental evolution in Escherichia coli. II. Changes in life-history traits during adaptation to a seasonal environment. American Naturalist.

Wernegreen, J.J., and Moran, N.A. (1999). Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. Mol. Biol. Evol. *16*, 83–97.

Zorzet, A., Pavlov, M.Y., Nilsson, A.I., Ehrenberg, M., and Andersson, D.I. (2010). Error-prone initiation factor 2 mutations reduce the fitness cost of antibiotic resistance. Mol. Microbiol. *75*, 1299–1313.

"That's the jolly old story"